A. M. Lee

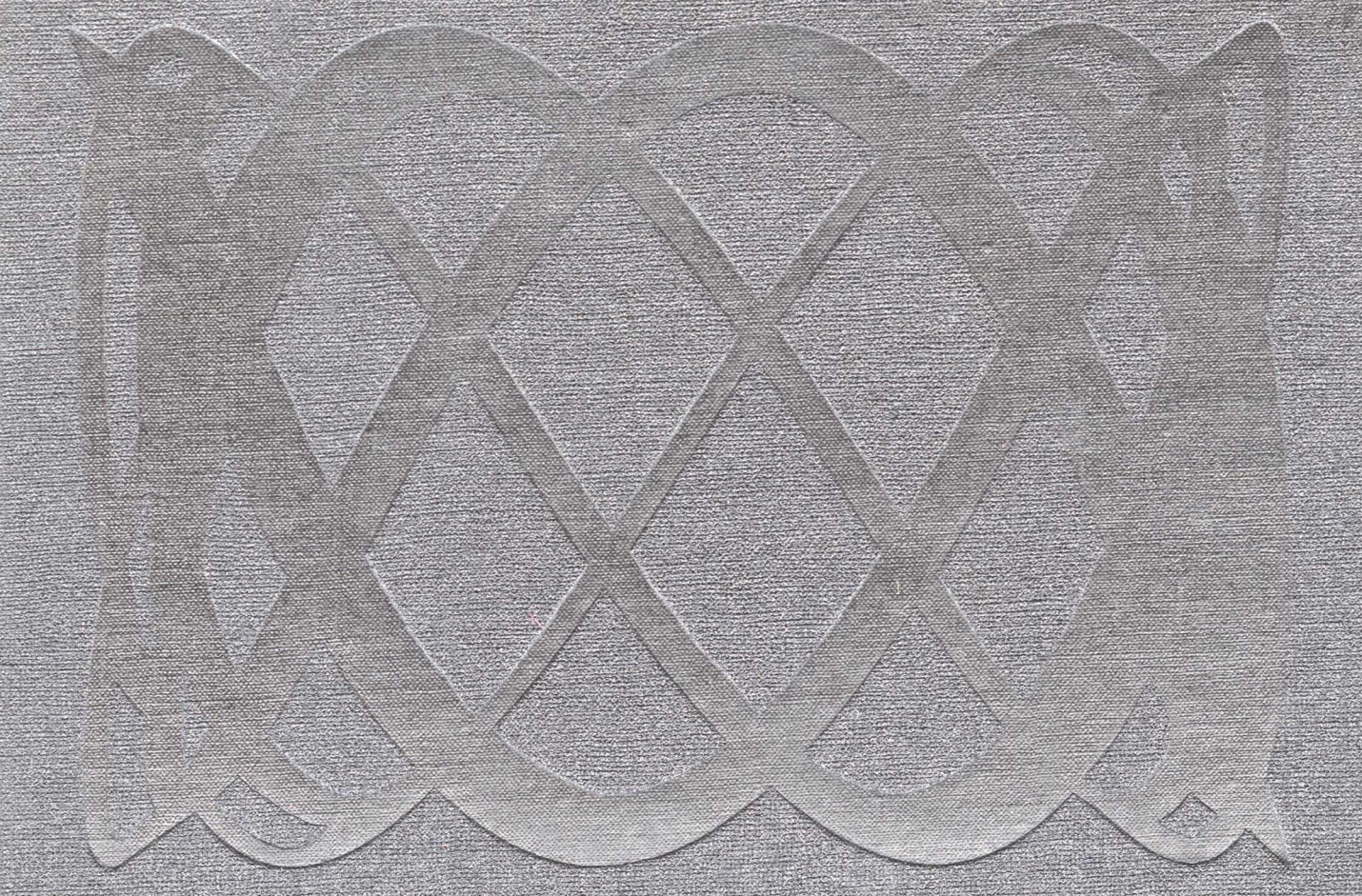
TEORIA AŞTEPTĂRII CU APLICAŢII

seria BAZELE MATEMATICE ALE CERCETĂRII OPERAȚIONALE

Principala problemă a aplicării teoriei așteptării constă nu atît în rezolvarea sistemelor de așteptare din lumea reală prin găsirea soluțiilor analitice ale unor modele foarte corecte, ci înainte de toate în găsirea unor soluții operaționale, aproximative, însă acceptabile și stabile.

Din acest punct de vedere, cartea lui A. M. Lee scoate în evidență anumite particularități ale sistemelor de așteptare din lumea reală, în care oameni servesc oameni ; se arată pe exemple, selectate în special din transporturile aeriene de călători, cum anumite ipoteze specifice unor modele clasice de așteptare nu se justifică în astfel de sisteme ; se prezintă în detaliu experiența autorului în soluționarea practică a unor astfel de sisteme.

Cartea este foarte utilă celor interesați : analiști și programatori, studenți ai institutelor tehnice și secțiilor de informatică, economiști și ingineri, arătînd pe multiple exemple cum se aplică unele modele teoretice în soluționarea unor probleme practice și prezentînd de asemenea, în detaliu, procedeul de simulare cu calculatorul a unor situații ce conduc la modele de așteptare complicate.



seria
BAZELE
MATEMATICE
ALE CERCETĂRII
OPERAȚIONALE

APPLIED QUEUEING THEORY

ALEC M. LEE, M.A.

Director of Operational Research, Air Canada

The Macmillan Press Limited

London

© Alec M. Lee 1966

Alec M. Lee

TEORIA AŞTEPTĂRII CU APLICAŢII

Traducere din limba engleză

de dr. VINCENŢIU DUMITRU, ANDREI ALEXIU şi TIBERIU SPIRCU

Editura tehnică

București — 1976

PREFAȚA LA EDIȚIA ROMÂNĂ

Cu toate că n-am vizitat România pînă în prezent, îmi imaginez că sistemele de așteptare sînt la fel ca oriunde; ca și canadienii, germanii sau japonezii, probabil că românii sînt la fel de iritați de aspectele neplăcute ale sistemelor de așteptare. De aceea, voi fi foarte mulțumit dacă cele scrise în cartea mea vor ajuta la reducerea inconvenientelor inerente sistemelor de așteptare.

În ultimii ani articolele și cărțile avînd ca temă teoria matematică a așteptării s-au înmulțit în mod alarmant. Totuși puține sînt acelea care au dovedit că au o utilitate practică; de aceea n-am considerat necesar să mai adaug referințe bibliografice. Trebuie însă să scot în evidență faptul că în anii care au trecut din momentul scrierii cărții, tehnicile de simulare au devenit mult mai importante și mai folositoare decît am presupus atunci. Aceasta se datorește doar în parte progresului acestor tehnici în sine, fiind mai mult o consecință a tendinței foarte favorabile a costului calculelor, ca urmare a progresului rapid al tehnologiei echipamentelor de calcul. România, care a făcut din proiectarea și fabricarea echipamentelor de calcul o componentă semnificativă a dezvoltării ei tehnologice, este bine plasată pentru a fi beneficiara acestor tendințe.

Ednaston, Derbyshire
Februarie 1976

ALEC LEE

PREFAȚĂ

Revistele de cercetare operațională din toată lumea continuă să fie pline cu lucrări legate de teoria firelor de așteptare, iar un cititor ocazional ar putea să-și formeze părerea că problemele acestui subiect au fost deja epuizate. Nu este însă așa. Majoritatea acestorlucrări își au originea, se pare, în zona crepusculară a cercetării academice. În ele sînt explorate curajos misterele modelelor simple ale celor mai familiare procese de așteptare și sînt obținute binecunoscutele proprietăți ale modelelor clasice În timp ce aparatul matematic devine din ce în ce mai bine pus la punct, rămîne într-adevăr o dificultate a găsi relatări ale muncii experimentale cu procesele de așteptare, sau a face observații empirice ale acestor procese, sau chiar aplicații ale unor teorii existente. Se pare că o mare parte a muncii de cercetare se depune pentru teorie și o mult mai mică parte pentru aplicații.

Există o aparentă inactivitate în aplicațiile teoriei așteptării, dar realitatea este alta. Problemele legate de aglomerație constituie un sector major în care este realizată foarte multă practică de cercetare operațională, par pentru a o descoperi este necesar — exceptînd sectoarele de transport și telecomunicații — *a asculta conversațiile de informare necontrolate, care au loc în relaxantele ore ce urmează conferințelor. Sînt mai multe motive pentru această reticență, dar unul este dominant în rîndurile multor analiști de cercetare operațională și anume că profesiunea lor este fără îndoială de natură matematică, cu un foarte ridicat nivel de rigoare. Ei știu că în propriile lor aplicații cotidiene ale teoriei așteptării asemenea rigoare nu este menținută, pentru că de obicei nu este posibil a formula modele ale proceselorde așteptare în termeni matematici foarte riguroși și a le rezolva. De aceea, acești analiști sînt refractari în a discuta experiențele lor pentru teama că ele vor fi expuse criticilor unor persoane competente. Dar această teamă nu este justificată. Mulți dintre cei mai competenți specialiști în teoria așteptării sînt intens preocupați de aplicațiile practice, chiar cînd aceste aplicații implică o mare ramificație și diversificare a modelelor teoretice. Mai mult, această atitudine de reticență îi separă pe studenții în cercetare operațională de realitățile fenomenelor pe care le analizează și duce la consolidarea ideii că modelele teoretice sînt altceva decît aproximări conceptuale ale realității.*

Lucrarea de față este o încercare de a abandona această reticență. Ea nu tratează despre teoria firelor de așteptare, ci despre problemele aplicării rezultatelor ei în practică. Sînt multe cărți adecvate, și chiar cîteva excelente, legate de această teorie și nu am avut niciodată nici interesul, nici competența să le completez numărul. Altfel spus, scopul acestei cărți este aplicarea teoriei așteptării în găsirea de soluții opera-

ționale problemelor de așteptare din lumea reală. Deoarece este dificil de găsit cazuri concrete în reviste (fac excepție unul sau două din articolele apărute mai recent cum ar fi „clasicul" „Traffic Delays and Toll-booths" al lui Edie), am fost nevoit să utilizez materiale din surse familiare sau de la prieteni, dintre care unele nu au fost pînă acum prezentate publicului larg. Aceasta a însemnat să folosesc material obținut din studii la care am fost eu însumi participant. Am urmat această cale cu regret. Însă pot fi sigur de faptul că toate cazurile prezentate sînt reale, ca și datele utilizate și — cu anumite excepții specificate în text — nefalsificate. Pericolele falsificărilor sînt limpezi.

Nu este o carte cuprinzătoare. Unele modele de așteptare ce au fost foarte mult studiate nu sînt menționate sau sînt prezentate doar în cîteva rînduri. Am punctat numai ceea ce cred necesar pentru atingerea scopului propus. Mai mult, nu există digresiuni legate de teoria probabilităților, statistica matematică sau de transformata Laplace. Se presupune că cititorul are anumite cunoștințe despre aceste subiecte. Orice cititor care este familiarizat cu vreuna dintre cărțile teoretice mai bune, legate de așteptare (cum ar fi cartea Elements of Queueing Theory *a lui Saaty) poate omite capitolele 2 — 5 ale prezentei cărți. Aceste capitole au fost scrise pentru cei ce se inițiază în acest domeniu sau pentru cei a căror memorie necesită o reîmprospătare printr-o recapitulare rapidă. Oricum, sper ca această carte luată global să fie interesantă și utilă.*

Sînt profund îndatorat multora pentru ajutorul acordat. O recunoaștere în acest sens este făcută în cadrul textului și notelor capitolelor. Trebuie oricum să-mi arăt recunoștința și să aduc mulțumirile mele Companiei British European Airways, precum și Companiei Air Canada pentru permisiunea de a utiliza extensiv materiale legate de studiile de cercetare operațională la care am participat în perioada 1956 — 1964. Fără acest ajutor n-aș fi putut nici măcar să încep să scriu această carte.

Aș dori de asemenea să mulțumesc secretarei mele, d-nei Jeannine Lacelle pentru magistrala transpunere a indescifrabilelor mele notițe; de asemenea soției mele, Barbara, pentru realizarea celor mai multe dintre grafice, a indexului și a unei mari cantități de cafea.

ALEC M. LEE

Montreal
Place Ville-Marie

1

INTRODUCERE

Pe măsură ce societăţile umane devin din ce în ce mai complexe, se pare că devine tot mai dificil pentru membrii lor de a evita participarea involuntară la procesele de aşteptare. Fiecare cititor al acestei cărţi trebuie să aibă o oarecare experienţă personală în acest domeniu. Pentru a cumpăra un timbru de la un oficiu poştal este necesar să stai la rînd pentru a fi servit. La aeroport, pentru a primi confirmarea unui zbor, de cele mai multe ori eşti nevoit să stai tot la rînd. Dimineaţa, înainte de a pleca de acasă, pentru a primi prin telefon date privind starea vremii, eşti obligat să formezi de mai multe ori numărul, pînă a obţine ceea ce doreşti. Toate acestea sînt procese de aşteptare.

Experienţa legată de aşteptare este foarte bogată. Dacă experienţa pasivă a unei activităţi pe parcursul unei perioade mai lungi ar duce la calificativul de expert (aşa cum mulţi practicieni insistă că este), atunci noi toţi am fi experţi în fenomene de aşteptare. De fapt, proporţia oamenilor ce pot fi găsiţi stînd la rînd la orice oficiu poştal şi care sînt experţi în materie de sisteme de aşteptare este redusă: şi s-ar putea admite, dar n-ar fi prea plăcut, că proporţia lor printre angajaţii oficiului poştal este încă şi mai mică. Nu poate fi presupusă prea multă familiaritate în legătură cu caracteristicile sistemelor de aşteptare. De aceea, ar fi nimerit a începe prin exemplificare, prin a stabili ceea ce înseamnă un sistem de aşteptare şi ce probleme se ridică cel mai semnificativ şi mai des în legătură cu asemenea sisteme. Deşi cele trei exemple ce urmează sînt descrise umoristic, ele sînt reprezentative pentru problemele şi procesele foarte reale. „Je me presse de rire de tout, de peur d'être obligé d'en pleurer". Ele prezintă întîmplări frecvente.

1.1. PROBLEMA OȚELĂRIILOR

Corporația „The Allied Coke and Iron" este un vast concern ce acoperă diverse activități economice. Printre componentele sale sînt mai multe întreprinderi de colectare a fierului vechi, precum și două oțelării. Cea mai veche oțelărie folosește procedeul Siemens de obținere a oțelului, avînd o nevoie constantă de fier vechi de bună calitate. Fierul vechi se transportă în camioanele întreprinderilor de colectare ale companiei. Camioanele sosesc la oțelărie și, înainte de a merge la furnale ca să-și descarce încărcătura, trec peste un cîntar-basculă, pentru a o cîntări și înregistra. Acesta este punctul critic.

La aproape orice oră, între 6 dimineața și 10 seara, la cîntarul-basculă este un șir (rînd) format din camioane ce așteaptă să fie cîntărite. Lungimea șirului variază ; dar aproape fiecare camion pierde o parte apreciabilă a timpului așteptînd. Unul dintre operatorii de la cîntarul-basculă are o idee : oare n-ar fi mai bine pentru companie să instaleze un al doilea cîntar-basculă și astfel să reducă timpul de așteptare al camioanelor? Oare costul instalării și folosirii cîntarului-basculă suplimentar nu va compensa posibilitatea de a spori utilizarea camioanelor?

Operatorul (pe nume Szymanowski) semnalează în scris sugestia sa și o înaintează conducerii companiei. După cîteva luni sugestia sa ajunge să fie examinată de către Comitetul director. Inginerul-șef adjunct este însărcinat să facă o evaluare sumară. El replică imediat că : „este clar că prin dublarea capacității de cîntărire timpul mediu de așteptare a fiecărui camion se va înjumătăți. Aceasta ne va aduce un cîștig relativ mic. Eu cred că propunerea serviciului tehnic, de a ne vinde camioanele și de a cumpăra unele mai mari, este mai valoroasă. Ea ne va costa doar 250 000 $. Ceilalți membri ai Comitetului director (puțin obosiți) sînt impresionați de acest raționament critic. Domnului Szymanowski i se mulțumește pentru sugestia sa „..., care, regretăm, este imposibil de materializat astăzi" ... și i se acordă o primă de consolare de 20 $.

1.2. PROBLEMA CASEI DE BILETE

Acțiunea are loc într-o gară din micuțul oraș scoțian Paltries, într-o frumoasă dimineață de iulie. Tocmai s-a oprit ploaia. Murdo

Macdonald intră să-și cumpere un bilet pînă la Edinburgh, pentru trenul de la ora nouă. El găsește deschis, ca de obicei, un singur ghișeu destinat vînzării biletelor de clasa întîi (pentru care nu este nici un amator) și un altul pentru clasa a doua (unde se află cincisprezece persoane aliniate așteptînd să le vină rîndul). Macdonald se așază la coadă pentru clasa a doua. El este un om rezonabil și, în ocaziile similare din trecut, a rezistat tentației de a reclama nedreptățile din organizarea căilor ferate. Chiar și astăzi el nu sugerează ca ambele case de bilete — cea pentru clasa întîi și cea pentru clasa a doua — să vîndă același gen de bilete și ca pasagerii pentru clasa întîi să fie constrînși astfel să stea la rînd cu ceilalți. El își dă seama că un pasager ce va călători cu clasa întîi plătește mai mult și este normal să i se ofere un serviciu superior calitativ, dar astăzi el este nespus de iritat de mișcarea înceată a funcționarului de la ghișeu și, așa cum se întîmplă de cele mai multe ori, din iritarea sa îi vine o idee. De ce să nu se vîndă biletele pentru ambele clase la amîndouă case de bilete, dar astfel încît să i se îngăduie fiecărui pasager ce dorește bilet pentru clasa întîi să intre în fața rîndului, de îndată ce a sosit? În acest fel privilegiile nu vor fi diminuate, în timp ce capacitatea de vînzare a biletelor va fi practic dublată (pentru că de fapt nu există pasageri pentru clasa întîi la Paltries). „Adoptînd această procedură" scria mai tîrziu Macdonald către editorul revistei „Paltries gazette and Stockbreeder's Tribune", „timpul mediu pierdut la rînd va fi redus în ceea ce mă privește, ca și în ceea ce privește și pe ceilalți pasageri de clasa a doua, și aceasta fără o cheltuială suplimentară a societății de căi ferate și fără vreun inconvenient pentru cei doritori de bilete de clasa întîi".

1.3. PROBLEMA SOȚIEI FURIOASE

William K. Overlander se găsește în barul clubului de golf Yellowridge, sorbind un Martini și adresîndu-se unui coleg de club. Este indignat. El tocmai a emis o instrucțiune în cadrul companiei sale, al cărei efect va fi că în viitor nici unul dintre angajați (dar absolut *nici unul*) nu va mai utiliza serviciile Companiei aeriene Transcental Airlines pentru călătoriile oficiale de afaceri; aceasta din cauză că în ziua anterioară soția lui a avut o experiență teribilă.

„După cum îţi spuneam", povesteşte el printre înghiţituri, „Marilyn a format de patru ori, în interval de 30 de minute, numărul de telefon al biroului lor de rezervări de bilete şi nu a reuşit să-l prindă. Apoi, în sfîrşit, a cincea oară a reuşit. Mi-a povestit că a sunat de trei ori şi apoi a răspuns o voce ce părea mai curînd o înregistrare pe bandă, care i-a spus că agenţii sînt foarte ocupaţi şi să aştepte puţin. Ei bine, a aşteptat circa jumătate de minut, timp în care nu s-a petrecut nimic, dar *absolut nimic*, cu excepţia intervenţiei vocii înregistrate ce a repetat mesajul anterior. Ca urmare, a închis".

„Ce a urmat?" — îl întreabă prietenul său. „A încercat din nou?"

„Nici vorbă", zise Overlander. „Aceasta nu a fost prima dată cînd se întîmplă aşa ceva la Transcedental. Nu, domnule, soţia mea a sunat la Compania aeriană Commonwealth şi Caribbean Airways şi ştii ce s-a întîmplat? A dat de ei de prima oară, fără nici un fel de aşteptare. Iată ceea ce se numeşte într-adevăr un bun serviciu. Băieţii de la Transcedental trebuie să capete noi ajutoare, după părerea mea. Probabil ei au numai o singură bandă înregistrată şi trei femei bătrîne la panoul de comandă."

Overlander, ale cărui sentimente le înţelegem şi le respectăm, ar fi surprins să afle că Transcedental are angajaţi exact acelaşi număr de agenţi pentru biroul de rezervări de bilete, ca şi Commonwealth şi Caribbean, pentru a servi (foarte convingător şi coincidenţial din punctul nostru de vedere) exact acelaşi număr de pasageri.

Nu este vorba despre un număr prea mic de funcţionari sau de un refuz de a investi bani; nu acestea sînt cauzele pentru care cei de la Transcedental sînt mai puţin prompţi în efectuarea serviciilor. Este vorba, după cum se va vedea, de ignoranţă.

1.4. NATURA SISTEMELOR DE AŞTEPTARE

Aceste trei povestioare moraliste sînt toate despre procese de aşteptare şi ilustrează modul în care se pun problemele legate de astfel de procese. În toate trei există *clienţi* care cer să fie *serviţi*. Camioanele cu fier vechi trebuie să fie cîntărite; cetăţeanul Macdonald vrea să i se vîndă un bilet de tren; stimabila cetăţeancă Overlander vrea să vorbească cu un agent al unei companii aviatice pentru a i se rezerva un bilet de avion.

Clienții doresc să fie serviți imediat, dar dorințele lor nu sînt satisfăcute imediat, ei trebuind să aștepte la rînd. Momentele în care ei se așază la rînd se numesc *momente de sosire*. Intervalul dintre momentul sosirii unui client și momentul în care el este servit se numește *timp de așteptare*. Clienții sînt serviți de către *stații de servire*, numite uneori și *canale* sau *ghișee*. Termenii : stație de servire, canal sau ghișeu semnifică același lucru și în această carte vor fi folosiți alternativ. Cîntarul-basculă din primul exemplu este un *ghișeu*; funcționarii de la ghișeu din al doilea exemplu reprezintă *stații de servire;* fiecare telefon de la biroul pentru rezervări de bilete pentru avioane este un *canal*. Fiecare termen își are propria valoare ilustrativă.

Timpul scurs pentru satisfacerea serviciului cerut de client se numește *timp de servire*. Suma dintre timpul de așteptare al clientului și timpul de servire este îndeobște cunoscută ca fiind *timpul total de așteptare*. De cele mai multe ori este mai puțin important să ne gîndim la timpii totali de așteptare, ci mai degrabă la timpii de așteptare la rînd, dar apar frecvent și excepții.

Tipurile de întrebări ce se pun de obicei, în legătură cu sistemele de așteptare, sînt următoarele :

a) Cîte stații de servire trebuie instalate suplimentar, dacă numărul clienților și timpul pentru servirea fiecărui client nu se schimbă, pentru *a reduce* procentul clienților care au de așteptat mai mult de 2 min la mai puțin de 5% ?

b) Ce efect are introducerea priorităților pentru anumite clase de clienți asupra timpului de așteptare?

c) Cît de mult trebuie mărit un fir de așteptare pentru a reduce pierderea clienților la un procent mai mic decît 10% ?

d) Ce ar fi mai convenabil : să se instaleze o altă stație de servire sau să se mărească camera de așteptare ?

e) Dacă există două clase de clienți, o anumită calitate a servirii clienților (măsurată prin timpul mediu de așteptare) poate fi realizată mai ieftin prin înzestrarea fiecărei clase cu cîte un grup separat de servire sau prin introducerea unui singur grup care să servească ambele clase ?

Acestea sînt tipurile de probleme abordate și tratate în cadrul acestei cărți.

2

MODELE ALE SISTEMELOR DE AŞTEPTARE

2.1. GENERALITĂŢI

Nu există o ştiinţă a sistemelor de aşteptare. Există o teorie a aşteptării legată de studiul unor modele matematice riguroase, dar care dă impresia că se bazează pe un surprinzător de şubred postament de cunoştinţe empirice. Nu sînt multe rezultatele experimentale obţinute în legătură cu procesele de aşteptare. Aşa după cum se va vedea, acest fapt bizar — din punct de vedere ştiinţific — are influenţe negative privind aplicabilitatea şi utilitatea rezultatelor obţinute pe baza teoriei. Dar înainte de a studia acest aspect deosebit de important, precum şi consecinţele sale, va trebui să dăm un răspuns la două întrebări preliminare: ce este un model simbolic — sau matematic*) şi care este scopul unui astfel de model? Pentru un model al sistemului de aşteptare din staţia de cale ferată Paltries putem considera următorii factori:

1. Particularităţile funcţionarului de serviciu (James Cameron este conştiincios, dar lent; Angus Sutherland este un tip iute, dar înclinat să dea bilete greşite sau să nu dea corect restul; Alan Robertson este rapid şi eficient; Jan Macrae nu este deloc practic).

2. Precizarea zilei din săptămînă (sînt mai mulţi pasageri de luni pînă vineri decît sîmbetele sau duminicile).

---

*) Churchman, C. W., Akcoff R., Arnoff, L. *Introduction to Operations Research*. New York, Wiley, 1957.

3. Starea vremii — dacă a fost foarte frig în timpul nopții sau nu (variația de temperatură de la Robbinston la Paltries este foarte mare și dacă roua a înghețat pe linia ferată, trenul va întîrzia ; nu va mai exista graba în acest caz în procesul de așteptare).

4. Prezența sau absența călătoarei Fiona McTaggart (dacă ea se va afla la coadă pentru un bilet, atunci operația va întîrzia îngrozitor de mult — pînă la disperare — pentru că : a) nu-și găsește niciodată portofelul pentru a-și putea plăti biletul ; b) parlamentează cu fiecare funcționar de la ghișeu despre o reducere de preț al biletului, care de fapt nu poate fi obținută decît joia în iulie, după ora 17).

5. Natura călătoriei pasagerilor. (Majoritatea pasagerilor merg la Edinburgh în partea de sus și la Robbiston în partea de jos, aceștia necreînd nici o problemă. Dar, întrucît nu există bilete tipărite pentru stațiile de după Berwick-on-Tweed, în cazul cînd un pasager dorește totuși un astfel de bilet, acesta trebuie scris de mînă pe un formular special, apoi înregistrat într-un registru negru. Aceasta, evident, necesită timp.)

6. Temperatura și gradul de umiditate existent. (Timpul de servire a unui pasager crește proporțional cu ambele elemente, exceptînd cazul în care funcționarul de serviciu este Alan Robertson care s-a născut la Madras și se simte în elementul său atunci cînd este cald.)

Pentru procesele de așteptare în care sînt implicați clienții-oameni și oameni ce realizează serviciul cerut, oricine care a avut un contact cu astfel de probleme un timp suficient de lung poate alcătui o listă asemănătoare, cuprinzînd un șir de factori de acest gen.

Evident, nu toți factorii au aceeași importanță. Cititorul va intui că numărul 4 — McTaggart — este cu mult mai puțin important decît numerele 1, 2 și 6. Dacă elementul esențial al procesului îl constituie aglomerația din orele de vîrf, atunci factorul 3 va fi semnificativ doar dacă ora de vîrf este în primele ore ale dimineții sau momentul de vîrf este iarna ; în celelalte cazuri nu. Raționînd în acest mod, cititorul poate scurta lista factorilor ce trebuie luați în considerare în orice moment al procesului.

Să ne oprim puțin asupra dezamăgirilor d-lui Karel Szymanowski. Factorii ce influențează operația de cîntărire pot fi următorii : 1° frecvența (intensitatea) sosirilor camioanelor ; 2° particularitățile operatorului de la cîntar ; 3° încărcarea medie a unui camion (camioanele excesiv încărcate se vor deplasa mai lent și vor pierde astfel mai mult timp în deplasarea în rînd spre cîntar) ; 4° camioanelor de

zece tone le este permis să se așeze la rînd în fața vagoanelor de cinci tone dacă și numai dacă ele transportă deșeuri. Etc.

După cum se poate remarca, acești factori nu sînt aceiași cu cei pe care i-am găsit în ilustrarea anterioară, nici nu putem presupune că ei interacționează în același mod. Totuși unii sînt comuni pentru ambele cazuri. Selectînd, reducînd și comparînd listele de factori care influențează funcționarea altor sisteme de așteptare, putem identifica un număr mic de factori comuni și de semnificație majoră în toate cazurile; este vorba de factori *primari*. Factorii care nu sînt comuni tuturor proceselor de așteptare, apărînd doar în anumite procese de așteptare sau chiar numai într-unul, sînt factori *secundari*. Vor trebui luate în considerare și legăturile și interdependențele dintre factori.

Se poate conchide că factorii primari sînt:

1° frecvența și modul de sosire al clienților;
2° repartiția timpilor de servire;
3° ordinea în care clienții sînt serviți;
4° numărul de stații de servire;
5° repartiția timpilor de așteptare ale clienților.

În general sîntem interesați în a afla efectele modificărilor factorilor 1 — 4 asupra factorului 5. Repartiția timpului de așteptare este o *măsură a eficacității* sistemelor de așteptare. (Vom vedea că mai există și alte măsuri semnificative și utile*). În plus, putem trage concluzia că legăturile semnificative dintre factori sînt:

1. Legătura dintre parametrii repartiției timpilor de servire și frecvența medie de sosire a clienților.

2. Legătura dintre numărul de fire de așteptare și intensitatea medie de sosire.

3. Legătura dintre disciplina de așteptare și numărul de canale în funcțiune.

Sîntem acum, într-o măsură mai mare sau mai mică, în situația de a începe formularea unui model general al sistemelor de așteptare.

2.2. CARACTERISTICILE MODELELOR DE AȘTEPTARE

Modelul matematic constă în cîteva legături simbolice între factorii primari. Prin atribuirea unor forme particulare repartițiilor

* Pentru detalii se poate consulta: Morse, P. M., Kimball, G. *Methods of Operations Research*. 1965.

statistice asociate acestor factori primari și prin atribuirea unor valori tuturor parametrilor incluși în model, vom putea realiza previziuni. Putem spune, de exemplu, care va fi media timpilor de așteptare a clienților în caz că ar fi zece stații de servire și care ar fi această medie în caz că ar fi 12, toate celelalte elemente rămînînd aceleași.

Obiectivul unui model matematic simbolic este de a realiza previziuni posibile despre comportarea sistemelor sub influența modificării condițiilor sau, uneori, despre sisteme care deocamdată nu există. De aceea, teoria așteptării este o teorie generală; este vorba nu de comportarea sistemelor de așteptare propriu-zise, ci a unor modele simbolice abstracte ale acestora.* Este o idee importantă, ce trebuie reținută.

Este ușor a cădea în greșeală crezînd că modelul ar fi însuși sistemul. Dezvoltarea teoriei așteptării, așa cum este indicată de natura principalelor modele studiate, pare a fi puternic influențată de două lucruri, anume de originea ei, origine ce se află în studiul aglomerației sistemelor telefonice, și de aceea ce este ușor și posibil în analiza matematică**. În cadrul acestui capitol vom observa că multe dintre presupunerile cel mai frecvent făcute, referitoare la studiile de așteptare, sînt tocmai acelea care par realiste și valide în realizarea și studierea modelelor proceselor de așteptare asociate sistemelor telefonice. Acestea însă nu pot fi aplicate în toate ocaziile; ele sînt neadecvate mai cu seamă atunci cînd clienții și stațiile de servire sînt oameni. Vom observa de asemenea că aceste presupuneri sînt chiar acelea care conduc la modele susceptibile de a le fi aplicată analiza matematică. (Matematicienii nu sînt mai dornici decît alții să-și dedice preocupările unor lucruri inabordabile.)

Vom analiza principalele caracteristici ale sistemelor de așteptare și în principal patru care sînt de cea mai mare utilitate în vederea unei clasificări. Acestea sînt: *fluxul de intrare*, *mecanismul de servire*, *disciplina de așteptare* și *numărul stațiilor de servire* (canalelor). Terminologia a fost creată de D. G. Kendall într-o carte celebră prezentată la Societatea Regală de Statistică din Londra,

---

* Tocher, K. D. *The Rôle of Models in Operational Research* J. Roy. Statist. Soc., Pt. 2, 1961.

** Pentru un scurt istoric care indică geneza acestei situații, vezi: Saaty (a), cap. 1, secțiunea 1—6. Cititorul poate încerca de asemenea și Sysky, R. *Introduction to Congestion Theory in Telephone Systems.* Edinburgh, Oliver and Boyd, 1960. Vezi în special primul capitol.

în 1951*. Această lucrare, poate mai mult decît altele, a avut ca urmare marele interes pentru teoria așteptării, care a preocupat Anglia după anul 1950. Kendall și-a extins nomenclatorul într-un sistem de clasificare care este astăzi universal.

Clasificarea lui Kendall este de forma : $A/B/s$, unde A specifică modul de intrare, B este repartiția timpilor de servire, iar s este numărul de stații de servire. În general se presupune că clienții sînt serviți în ordinea sosirii, deoarece codul clasificării nu permite definirea disciplinei de așteptare. Această schemă este însă prea vagă față de obiectivele acestei cărți. De aceea vom folosi o formă extinsă a notației lui Kendall, anume $A/B/s : (d/e)$, unde A, B, s au aceeași semnificație pe care le-a dat-o Kendall, d este numărul maxim de clienți ce pot fi cuprinși în sistem în orice moment (incluzînd pe cei ce sînt serviți), iar e identifică disciplina de așteptare.

Cîteva exemple vor lămuri aceste probleme.

1. $M/M/1 : (20/\text{FIFO})$ reprezintă un model de așteptare cu intrare aleatoare, timpi de servire aleatori, o stație de servire ; capacitatea maximă a sistemului este de 20 (adică în timp ce 19 persoane așteaptă și una este servită, toți noii veniți sînt refuzați și pleacă), iar disciplina de așteptare este primul venit, primul servit.

2. $M/E_k/5 : (\infty/\text{SIRO})$ înseamnă intrare aleatoare, timp de servire erlangian, cinci stații de servire, capacitate a firului de așteptare infinită (adică nici o intrare nu este refuzată), iar disciplina de așteptare aleatoare. Chiar și această notație nu reușește să clasifice complet modelele de așteptare, dar este ceva mai utilă pentru scopul nostru, decît cea folosită de către Kendall.

2.2.1. FLUXUL DE INTRARE

Apelul telefonic este un fenomen care nu rezultă din acțiunea comună a mai multor persoane, ci apare sub o formă individuală. De aceea nu este surprinzător faptul că în majoritatea lucrărilor despre sistemele de așteptare s-a presupus că sosirile sînt individuale. Aceasta și pentru a înlesni calculele matematice ; ar fi însă greșit —

* Kendall, D. G. *Some Problems in the Theory of Queues.* J. Roy. Statist. Soc., Ser. B, 13, No. 2, 1951. Este o remarcabilă lucrare, în care pentru prima oară Kendall sugerează utilizarea conceptului de lanțuri Markov.

orice motiv am avea s-o facem să considerăm sosirile ca fiind individuale, în unele cazuri în care este destul de clar că nu este deloc așa.

În anumite sisteme de așteptare, clienții nu sosesc singuri, ci în grupuri. De exemplu, pasagerii pentru cursele aeriene sosesc la aeroport în grupuri familiale, delegații de afaceri etc. Primul pas ce trebuie făcut întotdeauna în examinarea unui sistem de așteptare este de a stabili dacă sosirile sînt individuale sau în grup (*agregate*). Momentele de timp în care clienții individuali sau grupurile intră în sistemul de așteptare se numesc *momente de sosire*, iar intervalele de timp dintre două momente de sosire consecutive se numesc *timpii dintre sosiri*.

În studiile privind procesele de așteptare este necesar a culege, datele privitoare la timpii dintre sosiri într-o formă convenabilă pentru a putea da răspuns la două întrebări :

1. Timpii dintre sosiri sînt oare dependenți, există oare între ei anumite corelații?

2. Timpii dintre sosiri sînt oare corelați cu alte elemente ale sistemului de așteptare?

De cele mai multe ori se presupune că răspunsul la ambele întrebări este negativ. Cititorul circumspect va crede în mod consecvent că răspunsurile pot fi presupuse negative atunci cînd procesele de așteptare se referă la apelurile telefonice* și că răspunsurile negative conduc la o tratare matematică mai simplă. Cu siguranță că va avea dreptate.

Vom da un exemplu de timpi dintre sosiri care nu sînt independenți**. Zborurile necesitate de schimburile comerciale sînt planificate de conducerea aeroportului. Timpii de sosire sînt planificați cu mult timp înainte. În perioadele de vîrf ale zilei (la marile aeroporturi) sosirile sînt planificate la intervale egale. Dar avioanele sosesc mai devreme sau mai tîrziu. Momentul real al sosirii oricărui avion se găsește în apropierea momentului planificat. Dacă întinderea acestei repartiții a sosirilor este redusă în comparație cu intervalele de timp planificate între sosiri, nepunctualitatea poate fi neglijată în cele

---

* Fry, T. C. *Probability and its Engineering Uses.* Second Edition, New York, D. Van Nostrand Co. Inc., 1965.

** Există mai multe lucrări ce tratează problemele de acest tip. Vom da patru exemple: Bowen, E. G., Pearcey, T. *Delays in the Flow of Air Traffic.* J. Roy. Aero. Soc., 52 (1948), p. 447 ; Pearcey, T. *Delays in the Landing of Air Traffic.* J. Roy, Aero. Soc., 52 (1948), p. 450 ; Bell, E. G. *Operatorional Research into Air Traffic.* J. Roy. Aero. Soc., 53 (1949), p. 331 ; Galliher, H. P., Wheeler, R. C. *Non-stationary Queueing Probabilities for Landing Congestion of Aircraft.* În : Operat. Res. 6 (1958), No. 2.

mai multe dintre cazuri. Dacă însă este mai mare, sosirile pot fi considerate — din nou din rațiuni practice — aleatoare *. Însă în cazul intermediar, cînd timpii planificați dintre sosiri sînt de ordinul primelor două dispersii ale abaterilor momentelor sosirilor reale de la momentele planificate, sosirile nu pot fi tratate ca regulate sau ca aleatoare. Timpii consecutivi dintre sosiri sînt corelați între ei, corelație ce nu poate fi ignorată fără riscul unei erori.

Următorul pas ce trebuie făcut constă în stabilirea repartiției statistice a timpilor dintre sosiri și pentru aceasta trebuie mai întîi să ne întrebăm : poate fi considerată oare această repartiție ca fiind staționară în timp? Altfel spus, putem presupune oare că repartiția timpilor dintre sosiri nu-și modifică în timp proprietățile statistice? Oare probabilitatea unui interval cu durata de x minute, ales în mod aleator, este independentă de momentul zilei în care acest interval a fost măsurat? În studiile teoretice făcute se presupune de obicei că repartițiile timpilor dintre sosiri *sînt* staționare. În realitate asemenea cazuri sînt rare. Figura 2.1 reprezintă un model zilnic al convorbirilor telefonice dintr-un birou pentru rezervări de bilete de avioane. Este limpede pe acest exemplu că media timpilor dintre sosiri este variabilă în timp (modul de variație este aproape sistematic cinci zile pe săptămînă).

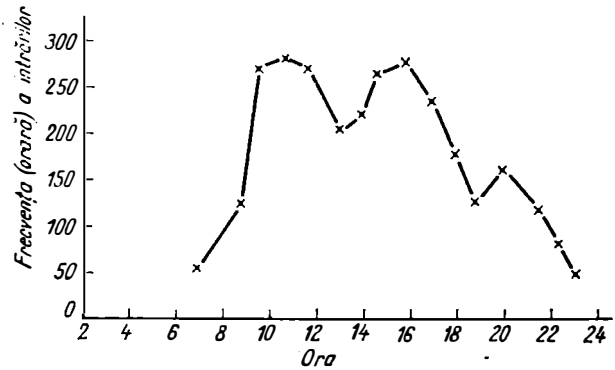


Fig. 2.1. Variația frecvențelor orare ale apelurilor telefonice într-o zi tipică (luna mai 1959).

* Cox și Smith, p. 14.

Această discrepanță dintre teorie și realitate nu este în fapt foarte mare. Majoritatea problemelor practice legate de procesele de așteptare sînt problemele *orelor de aglomerație*. Cu alte cuvinte, acestea devin foarte serioase cînd frecvența sosirilor este maximă. După cum vom vedea, pe parcursul unor asemenea perioade, dacă acestea nu sînt excesiv de lungi, repartițiile statistice ale timpilor dintre sosiri nu sînt de cele mai multe ori suficient de staționare pentru a ne putea permite să le considerăm astfel.

Forma repartiției timpilor dintre sosiri, care a fost cel mai mult studiată, este cea corespunzătoare așa-numitei repartiții a sosirilor *aleatoare*, adică funcției exponențiale negative. Densitatea de probabilitate a unui timp dintre sosiri de mărime t este în acest caz

$$a(t) = \lambda e^{-\lambda t},$$

unde λ este media sosirilor în unitatea de timp (unitate considerată în această carte a fi minutul, pentru cazul în care nu se specifică altceva). Aceasta este echivalent cu a spune că probabilitatea a n sosiri în unitatea de timp este dată de

$$e^{-\lambda} \frac{\lambda^n}{n!} \quad \text{(repartiția Poisson)}.$$

Semnificația acestei repartiții este că probabilitatea λdt a unei sosiri într-un interval mic de timp dt este constantă. Ea nu este afectată de momentul în care a avut loc ultima sosire, nici de numărul sosirilor care au avut deja loc. Într-o exprimare mai tehnică, intervalele dintre sosirile consecutive sînt variabile aleatoare independente între ele, toate avînd aceeași funcție de repartiție.

Cititorul își va reaminti că transformata Laplace a unei funcții exponențiale negative este

$$(1 - z/\lambda)^{-1}.$$

Funcția generatoare corespunzătoare repartiției Poisson este

$$e^{-\lambda(1-s)}.$$

Un alt tip de proces de intrare, care are o oarecare importanță practică, este cel numit de către Cox și Smith* *tipul cu intrare reînnoită*. Acesta poate avea loc atunci cînd intrările într-un sistem de așteptare sînt generate de ieșirile unui sistem anterior de așteptare, sau de emisii de la un număr finit de surse, ce emit fiecare cîte un client la intervale ce urmează o anumită repartiție. De exemplu, să considerăm k asemenea surse. A i-a sursă generează clienți la anumite momente, astfel încît intervalul dintre două generări succesive este repartizat cu o densitate de probabilitate $a_i(t)$. În acest caz fluxul combinat de sosiri format de ieșirile celor k surse are o repartiție specială a timpilor dintre sosiri. Este cunoscut faptul că în asemenea cazuri, forma statistică a sosirilor cumulate este aproximativ aleatoare pentru perioade mai scurte decît media intervalelor dintre sosirile de la toate sursele. În practică, intrările în sistemele de așteptare pot fi considerate ca fiind aleatoare.

2.2.2. MECANISMUL DE SERVIRE

Intervalul dintre momentul în care clientul începe să fie servit și momentul în care servirea este terminată este cunoscut sub denumirea de *timp de servire*. Este important să facem distincție între această mărime și *timpul de blocare* al stației de servire. Vom numi *timp de blocare* intervalul de timp dintre momentul în care începe timpul de servire al unui client și momentul în care începe timpul de servire pentru următorul client, în cazul că acest al doilea client a stat la coadă pentru a fi servit. Timpul de blocare poate fi conceput ca fiind timpul total în care un client împiedică servirea următorului, prin blocarea stației de servire.

Nu este necesar ca timpul de servire al unui client să coincidă cu timpul său de blocare. De exemplu, să considerăm un cititor care împrumută o carte de la bibliotecă. El sosește la ghișeu, cere cartea dorită, prezintă legitimația, bibliotecarul îi caută cartea, o înregistrează și i-o dă împreună cu legitimația. Din punctul de vedere al cititorului, *timpul* său de *servire* este încheiat și, ca atare, pleacă. Dar bibliotecarul nu servește imediat cititorul următor. El va pune înapoi fișa cărții și — dacă este cazul — o va înregsitra într-un dosar special (căci bibliotecarul Alderman Jones vrea să știe proporția

* Vezi comentariile făcute în Cox și Smith, pp. 106—107 și 141, ultimul paragraf.

cărților împrumutate din sectorul lucrărilor social-politice). Abia în acest moment el este gata pentru a servi următorul cititor : *timpul de blocare* s-a încheiat.

Diferența dintre timpul de servire și timpul de blocare poate apărea, în acest caz, banală. S-ar putea defini timpul de servire ca fiind chiar timpul de blocare, și astfel s-ar putea utiliza formulele standard. Este, fără îndoială, o considerație practică : munca realizată pe „timpul de servire" a clientului este, în general vorbind, obligatorie pentru ca clientul să elibereze stația. Cititorul nu va pleca fără cartea dorită. Restul servirii, efectuat în timpul de blocare, nu este neapărat necesar să fie realizat în acest timp, în scopul de a satisface clientul. Poate fi efectuat într-un alt moment al zilei. În acest fel, numărul celor ce vor sta la rînd va scădea simțitor. Metoda ar fi foarte eficientă dacă s-ar putea aplica tuturor sistemelor de servire, astfel încît, practic, timpii de servire să se confunde cu timpii de blocare. Următorul exemplu va reliefa semnificația distincției ce o facem între acești timpi.

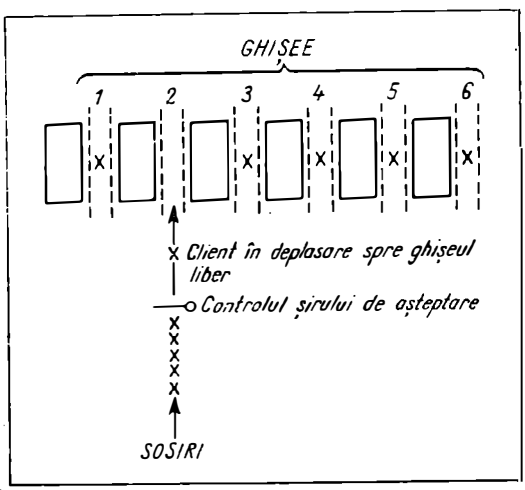


Fig. 2.2. Sistemul de așteptare al controlului pașapoartelor.

Procesul de așteptare ilustrat în fig. 2.2 nu este ieșit din comun ; el poate fi remarcat în operațiile privind controlul pașapoartelor, la multe aeroporturi internaționale.

Fie c ghișee de control al pașapoartelor, toate realizînd același tip de servire. Cînd un turist ajunge la unul dintre aceste ghișee, ofițerul de serviciu va începe executarea formalităților sau — dacă preferați — va începe „să-l servească". Timpii de servire pentru fiecare turist și pentru fiecare ghișeu în parte sînt independenți, identic repartizați. Toți turiștii sosiți se așază la coadă în ordinea strictă a sosirii, pe un singur șir. Nimănui nu i se permite să înainteze spre ghișeu, pînă cînd turistul anterior nu a fost servit. În sfera controlului pașapoartelor această regulă este de obicei controlată de un funcționar plasat în fața ghișeelor.

Dacă clientul ce urmează ar putea să se deplaseze cu viteza luminii la primul ghișeu liber, de îndată ce clientul anterior servit și-a încheiat timpul său de servire, atunci timpul de servire și timpul de blocare pentru acest client ar fi același. Aceasta este situația în sistemele telefonice, unde „clienții" sînt *convorbirile* care, evident, se pot deplasa cu viteză mare la o linie liberă. Dar pasagerii ce trec pe la ghișeele de control al pașapoartelor nu pot să se deplaseze cu viteza luminii. Ei merg cît mai calm și mai degajat; graba poate produce suspiciuni ofițerului de serviciu. Mai mult, distanțele din capul șirului pînă la ghișee nu sînt identice. Durează, în medie, mai mult a merge la ghișeele laterale decît la cel din centru. Putem presupune că există c variabile $x_i\,(i = 1,2, \ldots, c)$, reprezentînd timpii de deplasare pînă la fiecare din cele c ghișee. Timpul de blocare la ghișeul i este de aceea variabila v_i, unde $v_i = v + x_i$, v fiind timpul de servire al clientului.

În majoritatea analizelor proceselor de așteptare timpii de blocare sînt mai importanți decît timpii de servire și trebuie utilizați *în locul* timpilor de servire, orice formulă de așteptare ar fi fost folosită. Așa stînd lucrurile, procesul de așteptare de la controlul pașapoartelor nu este unul din cele mai simple; există mai multe stații de servire și fiecare are o altă repartiție a timpului de servire*. Sperăm că cititorul este convins de semnificația deosebirii dintre timpul de servire și timpul de blocare și va considera că ei sînt tratați, în majoritatea modelelor teoretice de așteptare, ca identici.

* Astfel de procese au fost studiate și modelate, dar nu sînt discutate în amănunt în această lucrare. Pentru unele detalii recomandăm lucrarea:
Gumbel, H. *Waiting Lines with Heterogeneous Servers*. În: Operat. Res., No. 4, 1960. Modelul lui Gumbel este aplicabil cu o disciplină de așteptare FIFO, preferabil în forma controlului total al intrărilor. În multe situații reale, în care există stații de servire eterogene (de exemplu ghișeele de la gări), nu există un singur șir controlabil, ci mai multe, între care clienții pot să se deplaseze.

Să mergem mai departe, atribuind deocamdată clienților calitățile obișnuite; vom presupune că timpul de servire și timpul de blocare sînt identici. Odată eliminată această dificultate, trecînd la examinarea proceselor de așteptare, trebuie să ne punem patru întrebări, legate de timpii de servire:

1. Oare toți clienții sînt de același tip, adică timpii lor de servire au aceeași funcție de repartiție?

2. Oare toate stațiile de servire sînt identice, adică timpii de servire legați de stațiile de servire au aceeași funcție de repartiție?

3. Timpii de servire sînt oare corelați cu alte aspecte ale sistemului?

4. Repartiția timpului de servire este oare staționară?

La începerea cercetării se speră întotdeauna că răspunsurile la aceste întrebări vor fi afirmative. În lucrările teoretice, de cele mai multe ori așa se și presupune; din păcate, însă, în practică aceste cazuri nu sînt prea întîlnite. Cîteva exemplificări o vor dovedi.

Cargouri ce transportă minereuri intră într-un port pentru a-și descărca încărcătura de minereu de fier. Ele au înzestrări și capacități diferite. De aceea, timpii necesari pentru a le descărca — timpii de servire — nu sînt omogeni din punct de vedere statistic. Ei au deci repartiții diferite.

În cadrul procesului de control al pașapoartelor timpul de blocare, care, în orice analiză a procesului, trebuie să înlocuiască timpul de servire, depinde de repartiții diferite ce corespund diferitelor ghișee.

Lingourile de oțel sosesc la laminoare de la cuptoare deschise. Ele sînt încă incandescente, dar prezintă variații de temperatură. Ele trebuie plasate în cuptoare de reîncălzire (puț infiltrat în pămînt) înainte de a fi laminate, operația de laminare necesitînd o temperatură înaltă a lingourilor. Dacă cuptoarele de reîncălzire sînt pline în momentul în care sosește un lingou, acesta trebuie să aștepte. În timpul acestei așteptări, el se răcește. Cu cît așteaptă mai mult, cu atît va trebui menținut mai mult în cuptorul de reîncălzire. Timpul de servire al acestei operații este timpul în care lingoul ocupă cuptorul de reîncălzire. De aceea, în acest caz, timpul de servire este direct proporțional cu timpul de așteptare, iar procesul este autoagravant.

Agenții liniilor aeriene civile lucrează opt ore pe schimb, la un aeroport, verificînd pasagerii pentru diferitele zboruri. Anumite studii legate de operațiile de verificare, făcute atît în Anglia cît și în S.U.A., au arătat că ele obosesc agenții și că timpii deservire sînt mai lungi și mai variabili spre sfîrșitul schimbului.

Apare astfel evident că răspunsurile afirmative de la primele pentru întrebări nu pot fi întîlnite întotdeauna în lumea reală. Deoarece teoria, în general, presupune contrariul, ne vom ocupa în continuare de anumite procese de așteptare pentru a reliefa cît de eronat poate fi un raționament care presupune afirmativ răspunsurile de mai sus și folosește modele teoretice gata făcute, acolo unde este clar că nu sînt adecvate. Desigur, cititorul va începe să creadă că, neputînd fi folosite în acest fel, multe modele de așteptare clasice familiare nu pot fi folosite la nimic altceva, decît în practica telecomunicațiilor. Deci întrebarea „Cît de temeinice sînt aceste modele?" nu este deloc de ignorat.

Acum, cînd unele aspecte au fost clarificate, vom trece sumar în revistă tipurile de repartiție a timpului de servire care sînt cel mai des utilizate și care au fost cel mai mult studiate. Vom începe cu repartiția exponențială negativă.

Repartiția exponențială negativă a mai fost considerată atunci cînd s-a vorbit despre intrările în sistemul de așteptare. În acest caz, densitatea de repartiție a timpului de servire *de durată* t este dată de relația

$$b(t) = \mu e^{-\mu t},$$

unde μ este numărul mediu de clienți serviți în unitatea de timp sau, altfel spus, $b = 1/\mu$ este timpul mediu de servire a unui client.

Din punct de vedere matematic, aceasta este cea mai convenabilă formă a repartiției timpului de servire. Explicația este că timpul rămas pînă la efectuarea deplină a unui serviciu este independent de timpul pentru care serviciul a fost în progres. Aceasta înseamnă, de exemplu, că dacă un client găsește doar o persoană înaintea sa și această persoană este în curs de servire, atunci timpul de așteptare este repartizat exact ca și timpul de servire. Acest lucru simplifică foarte mult analiza.

Repartiția exponențială negativă a timpului de servire nu este prea comună, dar nu sînt rare cazurile în care repartițiile timpilor de servire observați nu diferă prea mult de repartiția exponențială negativă. Cînd apare o deviație observabilă, este adesea posibil să o aproximăm cu repartiția Pearson de tipul III. Există suficiente motive matematice pentru alegerea unei repartiții de acest gen. De fapt, în general, este posibil a aproxima astfel de date cu *repartiția logaritmică normală* (vezi, de exemplu, cap. 15). Dar repartiția loga-

ritmică normală nu este tot atît de convenabil de manevrat, nici nu admite o interpretare mai convenabilă decît repartiția Pearson de tipul III.

Repartiția Pearson de tipul III, care este utilizată în modelele proceselor de așteptare, are densitatea de repartiție

$$\frac{1}{\Gamma(k)} (\mu k)^k \, e^{-\mu k t} t^{k-1}.$$

Atunci cînd k este un întreg, formula poate fi exprimată, în forma sa *erlangiană* (numită astfel după danezul Karl Erlang, cel care a utilizat această formă în studiile firelor de așteptare):

$$\frac{(\mu k)^k}{(k-1)!} e^{-\mu k t} t^{k-1}.$$

În această formă, repartiția este deci aceea a unei sume de k variabile independente, identic repartizate exponențial cu media $1/(\mu k)$. Aceasta este cea mai utilă legătură între repartiția Erlang și cea exponențială. Repartiția poate fi privită ca o modificare a repartiției χ^2 cu media $1/\mu$ și cu $2k$ grade de libertate.

Cititorul va remarca existența a doi parametri în formula repartiției, μ și k; astfel, ambii pot fi estimați utilizînd un set de observații. Repartiția este foarte asimetrică pentru valori mici ale lui k; pentru $k = 1$ este tocmai repartiția exponențială negativă. Pe măsură ce k crește, dispersia începe să descrească, pînă cînd, pentru un k suficient de mare repartiția devine aproape constantă.

Media repartiției este $1/\mu$, iar coeficientul de variație este $1/\sqrt{k}$. Transformata Laplace a repartiției este

$$[1 + z/\mu k]^{-k}.$$

Al treilea tip important de repartiție a timpului de servire este *repartiția constantă*. Aceasta însă nu necesită nici o prezentare.

2.2.3. DISCIPLINA DE AȘTEPTARE

În teoria așteptării se presupune ca ipoteză majoră asupra disciplinei de așteptare faptul că clienții sînt serviți în ordinea sosirii.

Această regulă este numită FIFO*, ceea ce înseamnă *„primul venit este primul servit"*. Disciplina de așteptare este elementul esențial de corectitudine din punctul de vedere al clienților, fiecare fiind preocupat de propriul său timp de așteptare. Ea nu este neapărat esențială din punctul de vedere al stației de servire și poate fi chiar nesemnificativă în situațiile în care clienții sînt indiferenți față de timpul cît așteaptă.

Merită să examinăm implicațiile regulii FIFO. În acest scop să considerăm sistemul de așteptare de la controlul pașapoartelor care a fost descris în secțiunea precedentă. Regula FIFO poate fi realizată doar într-un sistem de așteptare cu mai multe stații de servire, cînd :

1. Clienții sînt îndreptați, în ordinea sosirii lor, spre un singur șir de așteptare și rugați să se deplaseze mai departe spre capul șirului.

2. Un client poate părăsi șirul numai în clipa cînd a ajuns în față și i-a fost permis printr-un semnal să facă aceasta.

Ne permitem să comentăm aplicabilitatea acestei proceduri. Mai întîi, așa după cum remarcam și în capitolul precedent, regula FIFO poate fi foarte bine aplicată la sistemele telefonice, acolo unde clienții — în acest caz, convorbirile — sînt obligați, datorită circuitelor, „să înainteze pe o singură linie".

Regula poate fi utilizată, de asemenea, în condițiile clienților-oameni, prin adoptarea unei simple ordonări, întîlnită de obicei în magazinele alimentare nord-americane. Fiecare client, la sosire, primește un tichet numerotat. Atunci cînd un client a fost servit, vînzătorul va striga numărul următor. Persoana care are tichetul cu numărul menționat va înainta pentru a fi servită și va înmîna tichetul vînzătorului.

Regulile FIFO privind disciplina de așteptare sînt de fapt foarte larg răspîndite. Ele tind să apară ori de cîte ori :

1. Clienții sînt oameni sau fenomene abstracte asociate oamenilor — cum sînt convorbirile telefonice — care sînt foarte afectați de durata de așteptare sau de tratamentul diferențiat.

---

* Regula FIFO se referă la „primul venit, primul servit", aceasta fiind o formă mai exactă de descriere decît „primul sosit, primul plecat".

Într-un proces cu mai multe stații, primul client din stație nu o părăsește neapărat mai devreme decît alt client, care a intrat într-o stație diferită, la alt moment. În continuare, regulile disciplinei de așteptare, cum ar fi FIFO sau SIRO, se referă la modul în care clienții sînt selectați în fir pentru a fi serviți, neavînd nici o legătură cu ordinea în care ei părăsesc sistemul. În cazul sistemelor de așteptare cu un singur fir, evident că nu există nici o diferență atîta timp cît regulile priorităților relative nu sînt operaționale.

2. Există restricții de natură fizică (circuitele telefonice, numărul tichetelor) sau psihică (morală) (de exemplu atitudinea pasagerilor într-o stație de autobuz din Londra) care obligă să se înainteze în mod ordonat.

Ar fi însă greșit să considerăm că regulile FIFO ar fi universale, chiar atunci cînd clienții sînt oameni. De exemplu, să considerăm oamenii care așteaptă autobuzul într-o stație din Roma. Regula de aici ar putea fi comparată cu o altă regulă care a fost studiată destul de amănunțit : servirea în ordine aleatoare, cu alte cuvinte, regula SIRO. Regula SIRO a disciplinei de așteptare poate fi întîlnită în multe situații ca, de exemplu, în telefonie, în operațiile manuale efectuate la un tablou de comutație, la vînzarea răcoritoarelor la bufetul unui teatru, în pauzele piesei ; la rezervarea de cușete pentru pasageri. (Nu este o glumă referitoare la căile ferate ; de fapt, sînt foarte frecvente cazurile în care companiile feroviare nu pot comunica cu pasagerii înregistrați pentru rezervări, sunîndu-i la numărul de telefon pe care aceștia l-au lăsat, și vor continua căutările pînă cînd, printr-o șansă, vor găsi pe cineva care mai dorește o cușetă.)

Există anumite situații speciale, așa după cum vom vedea, în care regula LIFO (ultimul venit, primul servit) a disciplinei cozii ar fi foarte potrivită. De exemplu, în cazul lingourilor de oțel ce așteaptă să intre la reîncălzire, regula LIFO a disciplinei de așteptare ar putea fi cea mai eficientă. Prin utilizarea ei, căldura ar putea fi conservată, s-ar scurta timpul de servire și în cele din urmă s-ar reduce și timpul de așteptare. Se pare că așa este, ținînd seama de legea lui Stefan a răcirii.

Alte tipuri de disciplină a așteptării sînt asociate unor reguli de prioritate, atunci cînd există mai multe tipuri de clienți. Să presupunem că clienții fac parte din k clase și fiecare clasă prezintă o anumită prioritate. Clienții din clasa a i-a sosesc urmînd o repartiție dată de funcția $A_i(t)$ și sînt serviți, repartiția timpului de servire fiind dată de funcția $B_i(t)$. Clienții componenți ai unei aceleiași clase sînt serviți după regula FIFO, dar clienții cu prioritate i sînt serviți înaintea celor cu prioritatea $i+1$. Există două feluri de priorități. Dacă este în curs de servire un client cu prioritatea $i+1$, acesta va putea fi înlocuit și reîntors în șirul de așteptare în momentul cînd un client cu prioritate mai mare intră în sistem. Aceasta se numește *prioritate absolută*. Dacă totuși clientului în curs de servire i se poate termina serviciul, regula de prioritate urmînd să fie operativă după aceasta, avem de-a face cu un serviciu cu *prioritate relativă*. În cadrul unui model, vom

nota serviciul cu priorități absolute cu codul PSPO (ordinea de servire a priorităților absolute), iar prioritățile relative NPPS (ordinea de servire a priorităților relativă). Este destul de dificil să tratezi aceste priorități din punct de vedere matematic, exceptînd cazurile modelelor simple, și pînă în prezent nu li s-a acordat prea multă atenție. Acest aspect al priorităților ridică probleme interesante legate de cost, care nu a fost incorporat în modelele de așteptare așa cum ar fi trebuit. De exemplu, costurile pot fi introduse pentru a compara pierderile cauzate la neacordarea unei priorități mai mici clientului aflat în serviciu, cu pierderile totale sau parțiale rezultînd din eventuala lui eliminare.

În concluzie, se pot observa o mulțime de discipline de așteptare ce operează în multe sisteme reale de așteptare, dar este dificil a le formula matematic. De exemplu, există fenomenul deplasării*; deplasări individuale ale clienților de la un șir la altul, în scopul găsirii unui avantaj personal (aceasta în cazul cînd există un șir pentru fiecare din stațiile de servire, toate stațiile fiind identice). Pînă nu de mult, nimeni nu a dat prea multă atenție acestui fenomen, fie și de o manieră empirică.

Există, așadar, multe tipuri de discipline de așteptare, și pentru că regulile care guvernează ordinea servirii sînt printre aspectele foarte ușor controlabile și coordonabile ale proceselor de așteptare, ele prezintă un mare interes pentru cercetătorul operațional.

2.2.4. NUMĂRUL STAȚIILOR DE SERVIRE

Cînd într-un sistem de așteptare sînt mai multe stații disponibile pentru a realiza servirea, elementul esențial devine organizarea acesteia. Există două tipuri principale de organizare a servirii: *în paralel* și *în serie*. Prin stații de servire în paralel vom înțelege un număr de stații avînd facilități de servire identice, astfel încît un client să poată fi servit la fel de bine la oricare dintre ele. Cînd un client poate selecta sau poate fi îndrumat la oricare dintre stații, vom spune că stațiile sînt *complet abordabile*. Există, după cum vom vedea, posibilitatea ca să fie mai multe stații în funcțiune,

* Fenomen cunoscut sub denumirea de „jockeying". (*N.T.*)

dar care nu sînt complet abordabile, adică ele sînt complet abordabile doar de către anumiți clienți, dar nu și de către ceilalți. Cînd are loc acest lucru, vom vorbi despre *abordarea restrictivă.*

Stațiile de servire în serie se întîlnesc de obicei atunci cînd sistemul de servire este defalcat în mai multe elemente. Atunci, pentru a realiza servirea, se formează o serie de șiruri de așteptare, fiecare stație fiind destinată unui element al servirii. Clientul trece pe rînd pe la toate stațiile, începînd cu prima. El va avea de așteptat pentru a fi servit la toate fazele procesului.

Timpii de așteptare ai clientului și timpii totali de așteptare sînt puternic afectați de forma organizării și este important de a decide corect modul optim de alocare a capacității de servire. În general, în această carte, ne vom ocupa foarte sumar de sistemele de așteptare cu stații de servire în serie. Aceasta nu pentru că problema ar fi lipsită de importanță, ci datorită dificultății de a găsi cazuri potrivite pentru a le discuta.

3

MODELE CU O SINGURĂ STAȚIE DE SERVIRE

3.1. GENERALITĂȚI

Se obișnuiește să se înceapă expunerile în teoria așteptării cu o prezentare a diferitelor modele cu o singură stație de servire și în particular să se prezinte în detaliu aparatul matematic al celui mai cunoscut și mai utilizat dintre ele, adică al modelului $M/M/1$: (∞/FIFO). Este limpede că motivele pentru aceasta sînt în primul rînd de natură pedagogică. Însemnătatea practică a modelelor cu o stație de servire, de orice gen ar fi, este relativ neînsemnată. Atunci cînd se descoperă unele neajunsuri în funcționarea unui sistem cu o stație de servire, de obicei se încearcă să se evalueze posibilitățile transformării lui într-un sistem cu două stații de servire. Meritele pedagogice ale modelelor cu o stație de servire nu trebuie însă ignorate. Cu toată aplicabilitatea lor practică directă limitată, vom păstra tradiția și vom începe analiza cu ele.

Înainte de a trece la discutarea diferitelor modele cu o stație de servire, trebuie să menționăm că mulți autori au studiat modele generale de acest tip și „le-au rezolvat"*, din punct de vedere

---

* Considerăm că este util să prezentăm anumite exemple excelente din punct de vedere teoretic, cum ar fi :
Smith, W. L. *On the Distribution of Queueing Times.* În : Proc. Cambridge Phil. Soc., 49, 1953, pp. 449—461 ; Takacs, L. *Investigation of Waiting-time Problems by Production to Markov Processes.* În : Acta Math. Acad. Sci. Hung., 6, 1955, pp. 101—129. Din păcate lucrările prezintă un grad ridicat de matematizare.

matematic. Cu alte cuvinte, ei a considerat modele de tipul $GI/G/1$: (∞/FIFO) şi au exprimat repartiţiile timpilor de aşteptare şi a timpilor totali de aşteptare. De obicei, formulele obţinute corelează transformatele Laplace ale repartiţiilor timpilor de aşteptare şi a timpilor totali de aşteptare, cu tansformatele Laplace ale repartiţiilor timpilor dintre sosiri şi timpilor de servire. Ele nu sînt întotdeauna soluţii în sensul folosit de noi deoarece, cu excepţia formelor celor mai simple ale timpilor de aşteptare şi a timpilor de servire, transformatele Laplace nu pot fi inversate în mod efectiv. Este adevărat că sînt posibile evaluări numerice, dar de cele mai multe ori este foarte dificil a le obţine. Trebuie să menţionăm totuşi că majoritatea rezultatelor privitoare la modelele cu o staţie de servire ce vor fi discutate în acest capitol pot fi obţinute fără prea mari dificultăţi dintr-un model foarte general. Motivul pentru care nu vom proceda astfel constă în dorinţa de a prezenta şi alte tehnici ce sînt foarte utilizate.

3.2. MODELE CU TIMPI DE SERVIRE EXPONENŢIALI

3.2.1. $M/M/1$: (∞/FIFO)

Acesta este cel mai simplu model teoretic dintre cele ce au pretenţii de aplicabilitate. Vom reaminti caracteristicile lui, precum şi denumirile cîtorva parametri; astfel:

M Sosirile sînt aleatoare, dar au loc cu intensitate medie constantă (în această carte, unitatea de timp este minutul, în cazul în care nu este specificat altceva).

M Repartiţia timpului de servire este exponenţială negativă cu o valoare medie constantă, $1/\mu$ minute.

1 Există doar o staţie de servire.

∞ Nu există vreo limită legată de lungimea şirului, clienţii nu sînt respinşi de sistem, iar odată aşezaţi în şirul de aşteptare nu-l mai pot părăsi.

FIFO Clienţii sînt serviţi în ordinea strictă a sosirilor.

Modelul este uşor de mînuit, datorită unei ciudate proprietăţi a repartiţiei exponenţiale negative, ceea ce Saaty numeşte „uitarea". Prin aceasta ne referim la faptul că dacă timpii de servire au această repartiţie, atunci probabilitatea terminării serviciului la un moment dat este independentă de durata serviciului. Consecinţa acestei

proprietăți este că, stabilind ecuațiile pentru a releva modificările survenite în cadrul modelului, nu este necesar să ținem seamă de cît timp a trecut de la momentul ultimei sosiri, nici cît de mult a trecut pînă cînd a început timpul curent de servire.

Vom introduce acum simboluri (o listă completă a simbolurilor utilizate se găsește în Anexa 2). Probabilitatea de a avea o sosire pe parcursul unui interval de timp foarte scurt $(t, t + \mathrm{d}t)$ este $\lambda \mathrm{d}t$. În mod similar, dacă notăm $\mu = 1/b$, atunci probabilitatea ca unei persoane să i se termine servirea pe parcursul intervalului de timp ales (presupunînd că au existat încă de la început persoane în curs de servire) este $\mu\ \mathrm{d}t$. În fine, fie $p_n(t)$ probabilitatea ca la momentul t să fie n clienți în sistem. Probabilitatea $p_n(t + \mathrm{d}t)$ ca sistemul să conțină n persoane la momentul $t + \mathrm{d}t$ are trei componente :

— Probabilitatea $p_n(t)$ ca sistemul să conțină exact n persoane la momentul anterior t, înmulțită cu probabilitatea de a nu se înregistra nici o plecare sau sosire în intervalul de timp $\mathrm{d}t$ dintre aceste momente. Cum $\mathrm{d}t$ este foarte mic, această ultimă probabilitate este

$$1 - \lambda \mathrm{d}t - \mu \mathrm{d}t$$

iar contribuția totală a acestei componente este

$$[1 - (\lambda + \mu)\mathrm{d}t] \cdot p_n(t).$$

— Probabilitatea $p_{n-1}(t)$ ca sistemul să conțină exact $n-1$ persoane la momentul anterior t, înmulțită cu probabilitatea $\mu\ \mathrm{d}t$ a unei sosiri în intervalul $\mathrm{d}t$. (Cum $\mathrm{d}t$ este foarte mic, probabilitatea ca să existe mai mult de o sosire este neglijabilă.) Contribuția acestui termen este

$$\lambda \mathrm{d}t \cdot p_{n-1}(t).$$

— Probabilitatea ca sistemul să conțină $n + 1$ persoane la momentul precedent t, înmulțită cu probabilitatea $\mu \mathrm{d}t$ a unei plecări în intervalul $\mathrm{d}t$. (Probabilitatea mai multor plecări este neglijabilă.) Contribuția acestui termen este

$$\mu \mathrm{d}t \cdot p_{n+1}(t).$$

În consecință, forma generală a relației este

$$p_n(t + \mathrm{d}t) = [1 - (\lambda + \mu)\,\mathrm{d}t]\,p_n(t) + \lambda \mathrm{d}t\, p_{n-1}(t) + \mu \mathrm{d}t\, p_{n+1}(t),$$

și după o rearanjare a termenilor, împărțind prin $\mathrm{d}t$ și trecînd la limită pentru $\mathrm{d}t$ tinzînd către zero, rezultă că

$$\frac{\mathrm{d}}{\mathrm{d}t}\, p_n(t) = -(\lambda + \mu)\, p_n(t) + \lambda p_{n-1}(t) + \mu\, p_{n+1}(t) \qquad (3.1)$$

care este adevărată pentru $n \geqslant 1$.

Pentru $n = 0$, ceea ce înseamnă că stația este liberă (nu există nici o persoană care să aștepte și nici una care să fie servită), avem

$$\frac{\mathrm{d}}{\mathrm{d}t}\, p_0(t) = -\lambda\, p_0(t) + \mu p_1(t). \qquad (3.1\,\mathrm{a})$$

Aceste ecuații diferențiale constituie modelul matematic, care poate fi rezolvat, dacă se cunoaște numărul de persoane aflate la coadă la momentul inițial. Este mai convenabil să se ia momentul inițial $t = 0$, deși acest lucru nu este absolut obligatoriu. De exemplu, dacă ar fi m persoane la coadă la momentul 0, condițiile inițiale ar fi

$$p_m(0) = 1, \qquad p_n(0) = 0, \qquad n \neq m.$$

Cu această informație, ecuațiile scrise mai sus permit calcularea probabilităților $p_n(t)$ pentru orice t și pentru toate valorile $n = 0, 1, 2, \ldots$

Un caz important apare atunci cînd $m = 0$. Aceasta înseamnă că la momentul inițial nu există clienți în așteptare și deci stația de servire este liberă. Condițiile inițiale sînt

$$p_0(0) = 1, \quad p_n(0) = 0, \quad n > 0.$$

Cea mai avantajoasă metodă de obținere a unor relații explicite pentru $p_n(t)$ aparține lui Cox și Smith*. Această metodă învinge

* Cox și Smith, pp. 61—64: din nefericire rezultatul obținut este eronat, datorită unei erori de derivare. A fost corectată în ultima versiune.

dificultățile legate de faptul că ecuația (3.1 a) are o formă diferită față de cazul general, reprezentat de ecuațiile (3.1).

Înmulțind ecuațiile (3.1) cu s^n, unde s este o variabilă arbitrară, continuă și nenegativă, însumînd după valorile întregi ale lui n, $-\infty \leqslant n \leqslant \infty$, și introducînd funcția generatoare dependentă de timp $P(s, t)$, unde

$$P(s, t) = \sum_{n=-\infty}^{\infty} p_n(t)\, s^n,$$

atunci

$$\frac{\mathrm{d}}{\mathrm{d}t} P(s, t) = [-(\lambda + \mu)t + (\lambda s + \mu/s)t]. \tag{3.2}$$

$P(s, t)$ va furniza soluțiile dorite, dacă mai adăugăm condiția

$$p_0(t) = \frac{\lambda}{\mu} p_{-1}(t).$$

Soluția generală ce dă funcția generatoare a ecuației (3.2) este

$$P(s, t) = K(s) \cdot \exp\left\{-(\lambda + \mu)t + \left(\lambda s + \frac{\mu}{s}\right)t\right\}, \tag{3.3}$$

unde $K(s)$ este o funcție oarecare.

Termenul exponențial din partea dreaptă poate fi însă scris sub forma de doi factori; al doilea factor se scrie

$$\exp\left\{\left(\lambda s + \frac{\mu}{s}\right)t\right\} = \sum_{n=-\infty}^{\infty} I_n(2t\sqrt{\lambda\mu})\,(\lambda s/\mu)^{n/2},$$

unde $I_n(2t\sqrt{\lambda\mu})$ este o funcție Bessel de argument imaginar. Pentru prescurtare vom nota argumentul $2t\sqrt{\lambda\mu}$ cu x și vom reaminti că λ/μ este intensitatea traficului, notată de obicei cu ρ. Atunci

$$P(s, t) = K(s)\, \exp\{-(\lambda + \mu)t\} \sum_{n=-\infty}^{\infty} I_n(x)(\rho s)^{n/2}.$$

Presupunem că $K(s) = \sum_{r=-\infty}^{\infty} k_{-r} s^r$. Să înlocuim această expresie în relația de mai sus și să egalăm coeficienții puterilor lui s. Din condițiile inițiale $p_0(0) = 1$ și $p_n(0 = 0$ pentru $n > 0$ și proprietățile funcțiilor Bessel $[I_0(0) = 1$ și $I_n(0) = 0,\ n > 0]$, rezultă

$$k_0 = 1,\ k_{-r} = 0 \text{ pentru } r = 1, 2, \ldots$$

În consecință,

$$p_n(t) = \exp\{-(\lambda + \mu)\, t\}\left[\rho^{n/2} I_n(x) + \sum_{r=1}^{\infty} k_r\, \rho^{(n+r)/2} I_{n+r}(x)\right].$$

Valorile k_r sînt determinate astfel încît să asigure satisfacerea condiției $p_0(t) = \rho\, p_{-1}(t)$. Înlocuirea directă, urmată de egalarea coeficienților funcțiilor $I_r(x)$, arată că

$$k_1 = \rho^{-1},\quad k_2 = (1 - \rho)\,\rho^{-1},\quad k_3 = (1 - \rho)\rho^{-2} \text{ etc.}$$

Ca urmare, în final vom obține probabilitățile dependente de timp, de forma

$$p_n(t) = \exp\{-(\lambda + \mu)t\}\times$$

$$\times\left[\rho^{n/2} I_n(x) + \rho^{(n-1)/2} I_{n+1}(x) + \rho^n(1 - \rho)\sum_{r=2}^{\infty} \rho^{-(n+r)/2} I_{n+r}(x)\right] \quad (3.4)$$

Relația este adevărată pentru toți ρ nenegativi (chiar și pentru $\rho > 1$).

Pus în fața acestei relații, cititorul, avînd în minte ideea că acest sistem este cel mai simplu dintre toate, va avea poate un sentiment de panică. Acest sentiment va crește cu siguranță dacă va ieși la iveală faptul că soluția dată descrie doar creșterea șirului de așteptare (exprimată în număr de clienți ce așteaptă), nespunîndu-ne nimic despre timpii de așteptare a clienților. Pentru a-i obține și pe aceștia va trebui să considerăm expresii de formă mult mai complicată.

Dacă acestea sînt dificultățile găsirii soluției pentru modelele cele mai simple, ce se va întîmpla cu modelele mai complicate? Răspunsul la această întrebare este dat de prima regulă de lucru din teoria așteptării : soluțiile dependente de timp pentru modelele de așteptare sînt sau imposibil de obținut sau imposibil de manipulat.

Este însă liniștitor să constatăm că ocaziile în care se apelează la cercetători operaționali pentru a analiza sisteme de așteptare de acest fel sînt foarte rare. De asemenea, sînt rare cazurile în care se cere să se studieze un proces tocmai în decursul modificării prametrilor de sosire sau de servire. De obicei, nu se cere să se facă o analiză a sistemelor de așteptare pînă nu s-au descoperit dificultățile procesului, și de aceea, prin definiție, a trecut deja o perioadă de timp. Lăsînd deoparte acest mod simplist de a privi lucrurile, trebuie să constatăm că de fapt cea mai mare parte a problemelor practice legate de sistemele de așteptare nu prea au tangență cu comportarea tranzițională. În cazul în care stațiile de servire și/sau clienții sînt oameni, perioadele de început sînt marcate de o asemenea instabilitate a repartițiilor timpilor de servire, încît este practic imposibil să se apeleze la analiza matematică. În mod frecvent, repartițiile timpilor dintre sosiri sînt și ele instabile în aceste condiții. Ar fi deci absurd să cerem să aplicăm o soluție exactă a modelului în astfel de circumstanțe. De aceea, în cele ce urmează, exceptînd un singur caz, nu vom mai acorda atenție soluțiilor dependente de timp și formelor riguroase ale comportării tranzitionale. Iar atunci cînd va fi necesar să studiem astfel de probleme, vom utiliza metoda de aproximare sau tehnici de simulare.

Problemele majore ce se pun aici sînt de forma : oare cîte stații de servire sînt necesare, presupunînd că ele lucrează cu o intensitate normală? Intensități normale de obicei nu apar deloc în perioadele de început. Aceasta, ca și dificultățile matematice, explică de ce cea mai mare parte a studiilor făcute asupra sistemelor de așteptare au fost dedicate sistemelor ce se află în echilibru statistic*.

Un sistem de așteptare se află în *echilibru statistic* dacă probabilitatea existenței a n persoane în sistem (pentru orice $n \geqslant 0$) este independentă de timpul ce a trecut de la începerea procesului. Cu alte cuvinte :

$$p_n(t) = p_n.$$

Aceasta nu este, poate, cea mai elegantă definiție, dar ne va fi suficientă. Cititorul va observa că o altă formă de a exprima cele de mai

* Echilibrul statistic, sau starea stabilă, sau condițiile staționare, sînt riguros prezentate în Feller, pp. 356.

sus este de a spune că în echilibrul statistic (sau de *starea stabilă*) avem

$$\frac{\mathrm{d}}{\mathrm{d}t} p_n(t) = 0, \quad n \geqslant 0.$$

Astfel, în cazul nostru simplu, modelul se reduce la

$$\lambda p_0 = \mu p_1, \tag{3.5}$$

$$(\lambda + \mu) p_n = \lambda p_{n-1} + \mu p_{n+1}; \quad n \geqslant 1. \tag{3.6}$$

Aceste ecuații pot fi ușor rezolvate, dînd :

$$p_n = (1 - \rho)\, \rho^n, \quad n = 0, 1, \ldots, \tag{3.7}$$

unde $\rho = \lambda/\mu$ este *intensitatea traficului* măsurată în unități internaționale de măsurare a traficului numite *erlangi**. Repartiția lungimii șirului de așteptare este geometrică. Este limpede că aceasta este o funcție de repartiție numai în cazul în care ρ este mai mic decît 1. Cu alte cuvinte, echilibrul statistic există numai în condițiile în care intensitatea traficului este mai mică decît 1, adică atunci cînd stația de servire poate, în medie, servi clienții mai repede decît sosesc ei. Aceste aspecte evidente nu necesită o justificare matematică. De asemenea, este clar că dacă clienții sosesc mai repede decît poate să-i servească stația, șirul se va lungi continuu, iar probabilitatea ca șirul de așteptare să aibă o lungime dată va depinde de timpul scurs. Nu este atît de evident că în cazul în care intensitatea sosirii și intensitatea de servire sînt egale, adică atunci cînd $\rho = 1$, echilibrul statistic este din nou imposibil. Dar, în cazul în care $\rho = 1$, în expresia de mai sus p_n sînt toți egali cu zero și, fără îndoială, nu dau o funcție de repartiție.

Următoarea noastră sarcină este de a exprima repartițiile timpului total de așteptare și a timpului de așteptare în șir.

Transformata Laplace a repartiției timpului de servire este

$$\mu \int_0^\infty \mathrm{e}^{-\mu t}\, \mathrm{e}^{-zt}\, \mathrm{d}t = \frac{\mu}{\mu + z}.$$

* Vezi Syski (cap. 2, nota 4), pp. 13—14 pentru o analiză a măsurii traficului.

În consecință, transformata Laplace a sumei de r astfel de repartiții este $\left(\frac{\mu}{\mu + z}\right)^r$.

Ca urmare, transformata Laplace a timpului total de așteptare este

$$\sum_{r=0}^{\infty} p_r \left(\frac{\mu}{\mu + z}\right)^{r+1} = \sum_{r=0}^{\infty} (1 - \rho)\rho^r \left(\frac{\mu}{\mu + z}\right)^{r+1} = \frac{\mu(1 - \rho)}{\mu(1 - \rho) + z},$$

însă aceasta este tocmai transformata Laplace a expresiei

$$(\mu - \lambda)e^{-(\mu-\lambda)t}, \tag{3.8}$$

care este *densitatea de repartiție a timpului total de așteptare.*

Timpul mediu de așteptare este obținut prin diferențierea transformatei Laplace și făcînd apoi $z = 0$ din cauză că transformata Laplace este funcția generatoare a momentelor ; astfel :

$$W = \frac{1}{\mu(1 - \rho)}.$$

În mod similar se obține repartiția timpilor de așteptare în șir. Ea are valoarea $1 - \rho$ în 0 și, pentru $t > 0$, este dată de densitatea de repartiție $\rho(\lambda - \mu)\, e^{-(\lambda-\mu)t}$.

3.2.2. REPARTIȚIA PERIOADEI OCUPATE

Perioada ocupată este, în terminologia matematică a așteptării, intervalul de timp care începe atunci cînd sosește un client care nu găsește pe nimeni în fața lui și se sfîrșește cu prima ocazie cînd, la plecarea unui client, acesta nu lasă pe nimeni în urma sa. Nu este deloc același lucru cu *„ora de aglomerație"* care semnifică perioada în care intensitatea sosirilor este foarte ridicată. Această deosebire trebuie reținută, deoarece expresia *perioadă ocupată* va fi utilizată din cînd în cînd și în acest ultim sens.

Care este repartiția perioadelor ocupate? Aproape că am și determinat-o. Este dată de $(\mathrm{d}/\mathrm{d}t)p_0(t)$, căci aceasta măsoară frecvența de apariție a unei situații fără așteptare. Știm că

$$\frac{\mathrm{d}}{\mathrm{d}t} p_0(t) = -\lambda p_0(t) + \mu p_1(t),$$

iar pe $p_0(t)$ și $p_1(t)$ i-am obținut deja. Se poate arăta imediat că $\frac{\mathrm{d}}{\mathrm{d}t} p_0(t) = \lambda[I_2(x) - I_1(x)]$. Utilizînd următoarea proprietate a funcțiilor Bessel:

$$\frac{2n}{x} I_n(x) = I_{n+1}(x) - I_{n-1}(x),$$

formula de mai sus se reduce. Astfel, densitatea de repartiție a perioadelor ocupate devine

$$\frac{\exp\{-(\lambda + \mu)t\} \cdot I_1(2t\sqrt{\lambda\mu})}{t\sqrt{\rho}}. \tag{3.9}$$

Fără îndoială că aceasta este o expresie suficient de sofisticată dar nu neapărat imposibil de manipulat. Ea este însă destul de complicată pentru a ne sugera faptul că derivata repartiției perioadelor ocupate nu se obține prea ușor pentru modelele mai complicate. De asemenea nici formulele nu par a fi foarte ușor de manipulat. Poate cititorul se va întreba, în acest moment, care este importanța cunoașterii repartiției perioadei ocupate. Ca să fim sinceri, repartiția perioadei ocupate nu prezintă un interes deosebit. Vom vedea însă că unul dintre rezultatele intermediare care s-au obținut va fi deosebit de util.

3.2.3. $M/M/1 : (\infty/\mathrm{SIRO})$

Analiza acestui model este asemănătoare celui anterior, pînă în momentul în care trebuie determinată repartiția timpului de așteptare. Aceasta este evident, deoarece în deducerea repartiției mărimii șirului de așteptare nu s-a făcut nici o mențiune legată de disciplina

de așteptare. În orice caz, deducerea repartiției timpului de așteptare din repartiția mărimii șirului de așteptare, în cazul disciplinei SIRO, nu este o chestiune ușoară. Pentru a o face, este necesară utilizarea unui nou simbol :

$q_n(t)$ — probabilitatea ca un client, care atunci cînd sosește găsește deja n alți clienți în așteptare, este încă în șir și după ce a trecut un interval de timp t.

Să presupunem că clientul nostru sosește la un moment dt în loc de 0 și să considerăm următoarele evenimente ce au putut avea loc în acel scurt interval de timp ; găsim astfel :

Evenimentul (în intervalul dt)	Probabilitatea	Probabilitatea de așteptare în șir
1. Sosirea unui alt client	$\lambda\, dt$	$q_{n+1}(t - dt)$
2. Unul din ceilalți clienți începe să fie servit, după plecarea clientului precedent care a fost deja servit	$\frac{\mu n}{n+1}\, dt$	$q_{n-1}(t - dt)$ (Notă : la momentul dt în șir se află clientul nostru și alți $n-1$ clienți)
3. Nici o plecare și nici o sosire	$[1 - (\lambda + \mu)\, dt]$	$q_n(t - dt)$

Aceste elemente se combină între ele, ca în exemplul anterior ; după împărțirea prin dt și trecerea la limită se obțin următoarele ecuații de dependență a lui q_n de timp :

$$\frac{d}{dt} q_n(t) = -\frac{n}{n+1} \mu q_{n-1}(t) + \lambda q_{n+1}(t) - (\mu + \lambda)\, q_n(t), \quad n \geqslant 1, \quad (3.10)$$

$$\frac{d}{dt} q_0(t) = \lambda q_1(t) - (\mu + \lambda)\, q_0(t). \quad (3.11)$$

Condițiile inițiale sînt $q_n(0) = 1$ pentru toate valorile lui n, deoarece oricît de devreme s-ar așeza clientul nostru în șir, este sigur că el va sta acolo cel puțin un interval de timp zero.

Vom presupune de asemenea că intensitatea medie de servire este superioară intensității medii a sosirilor, adică $\rho < 1$, astfel încît clientul nostru va fi sigur că — mai devreme sau mai tîrziu — va fi

servit (se poate presupune că aceasta se va întîmpla mai tîrziu) și că $\lim_{t\to\infty} q_n(t) = 0$.

Problema este de a calcula $q_n(t)$ și acesta este momentul în care se pierde entuziasmul inițial. Nu vom schimba tradiția făcînd aceste calcule. Totuși, să presupunem valorile $q_n(t)$ cunoscute, în care caz probabilitatea ca orice client să aștepte cel puțin o perioadă t este

$$G_q(t) = (1 - \rho) \sum_{n=0}^{\infty} \rho^{n+1} q_n(t).$$

Am ținut seamă de faptul că repartiția mărimii șirului este independentă de disciplina de așteptare.

Nu este ușor de a face o analiză mai detaliată utilizînd pe $G_q(t)$. O altă cale este de a considera funcția $Q(t)$, probabilitatea unei întîrzieri de durată cel puțin t în servirea unui client deja întîrziat : aceasta exclude din calcul acei clienți care sînt serviți imediat.

Această expresie poate fi extinsă în serie de puteri ale lui μt, prin derivarea repetată a ecuațiilor (3.10) și (3.11). Această procedură a fost descrisă de Palm *. În acest mod, se poate arăta că timpul mediu de așteptare în șir este

$$W_q = \frac{\rho}{\mu(1 - \rho)}.$$

Este exact același ca și în cazul modelului bazat pe regula FIFO. Disciplina șirului nu afectează media timpului de așteptare (sau media timpului total de așteptare).

Cea mai utilă aproximare ** a probabilității $Q(t)$ ca o unitate din șir să aștepte mai mult de t intervale de timp o constituie aproximarea lui Riordan :

$$Q(t) \approx \frac{1}{2} \exp\{-\alpha(\mu - \lambda)t\} + \frac{1}{2} \beta \exp\{-\beta(\mu - \lambda)t\}, \quad (3.12)$$

unde

$$\alpha = 1 + \sqrt{\frac{\rho}{2}}, \quad \beta = 1 - \sqrt{\frac{\rho}{2}}.$$

---

* Vezi Palm, C. *Research on Telephone Traffic Carried by full-availability Groups.* TELE, English Edn, 1957.

** Riordan, J. *Delay Curves for Calls Served at Random.* În : Bell System Tech. J., 32, 1953, pp. 100—119, vezi anexa 3. Graficul A2—7, deși reprodus dintr-o altă sursă, este elementul fundamental al lui Riordan, cuprinzînd rezultatele cercetărilor efectuate de acesta.

3.2.4. $M/M/1 : (\infty/\text{LIFO})$

Funcția de repartiție a numărului de clienți în șir este în acest caz similară celei din cazurile precedente. Aceasta înseamnă că, în starea de echilibru, în particular :

$$p_n = (1 - \rho)\rho^n.$$

Nu același lucru se poate spune și despre repartiția timpului total de așteptare. Dacă putem considera regula FIFO a disciplinei de așteptare ca fiind cea mai echitabilă pentru clienți, atunci cea mai neechitabilă este regula LIFO : ultimul venit, primul servit. Care *este* în acest caz repartiția timpului total de așteptare? Răspunsul la această întrebare necesită o analiză foarte detaliată. Să presupunem că un client sosește și se așază în șir. Timpul lui de așteptare nu va fi afectat de numărul persoanelor existente în așteptare, deoarece conform regulii LIFO, el va fi servit înaintea celorlalți. Timpul lui va depinde numai de numărul clienților care sosesc *după* el și *înainte* ca el să înceapă să fie servit.

Cititorul își va reaminti repartiția perioadelor ocupate, pe care am dedus-o pentru modelul $M/M/1 :(\infty/\text{FIFO})$; o scurtă reflecție îl va convinge că e aceeași cu repartiția timpului total de așteptare în cazul modelului pe care-l analizăm. Să urmărim raționamentul pe care se bazează această afirmație. Un client — *subiectul* — sosind, găsește n persoane așteptînd, incluzîndu-l pe cel care este servit. El se duce în capul șirului. Pe perioada care trece de la sosirea sa și pînă cînd va fi servit, pot sosi și alți clienți ; în acest caz, ei îl vor preceda în șir. Cu alte cuvinte, subiectul va începe să fie servit în momentul în care un client care pleacă lasă tot n persoane în șir. Deci repartiția timpului total de așteptare în modelul $M/M/1 :(\infty/\text{LIFO})$ este repartiția „perioadei ocupate", începînd cu momentul în care se găsesc n persoane în șir și terminînd cînd numărul celor din șir este din nou egal cu n. Tranzițiile de acest fel sînt independente de n, așa încît perioadele ocupate sînt aceleași pentru regulile FIFO și LIFO. Cititorul va observa că dacă timpul de servire este repartizat exponențial, atunci timpul care trece între momentul sosirii subiectului și momentul servirii lui este de asemenea repartizat exponențial. Prin urmare, momentul sosirii subiectului nu are importanță ; timpul de așteptare este exact același cu al întregii perioade ocupate, generate de un timp complet de servire.

De aceea, densitatea de repartiție a timpului total de așteptare este

$$\frac{\exp\{-(\lambda + \mu)t\} \cdot I_1(2t\sqrt{\lambda\mu})}{t\sqrt{\rho}}. \tag{3.13}$$

Cititorul va putea arăta direct că durata medie a unei perioade ocupate este $1/[\mu(1 - \rho)]$, astfel încît timpul mediu de așteptare în modelul nostru este dat de

$$W = \frac{1}{\mu(1 - \rho)}. \tag{3.14}$$

Să subliniem din nou faptul că timpul mediu de așteptare într-un model $M/M/1$ este independent de disciplina de așteptare.

Cox și Smith* au realizat o deducere foarte ingenioasă a duratei medii a perioadelor ocupate. Timpul de lucru al stației de servire alternează între perioade ocupate și libere. Cînd sosirile sînt aleatoare, o perioadă liberă are șansa λdt de a se sfîrși într-un interval scurt dt. Urmează deci că durata perioadelor libere este repartizată exponențial cu media $1/\lambda$ sau a. Independent de natura repartiției timpului de servire, probabilitatea ca stația să fie liberă este $1 - \rho$. Considerînd o perioadă foarte lungă, T, stația este liberă pentru durata $T(1 - \rho)$ și în acest timp există $\lambda T(1 - \rho)$ perioade libere. De asemenea, trebuie să fie $\lambda T(1 - \rho)$ perioade ocupate, însă pentru o durată $T\rho$. Prin urmare, durata medie a unei perioade ocupate este

$$\frac{T}{\lambda T(1 - \rho)} = \frac{1}{\mu(1 - \rho)}.$$

Astfel, am ajuns din nou la timpul mediu de așteptare pentru modelul nostru.

3.3. MODELE CU TIMPI DE SERVIRE NEEXPONENȚIALI

Concluziile trase în secțiunea precedentă ne conduc la studiul modelelor cu o stație de servire, în care repartiția timpului de servire are altă formă decît cea exponențială. Exceptînd două cazuri, nu

* Cox și Smith, p. 58.

vom îngreuna expunerea prin examinarea deducerii rezultatelor pentru modele de acest gen, căutînd doar să rezumăm aceste rezultate. Excepțiile sînt menite a ilustra două feluri de argumente, utile în dezvoltarea ulterioară a modelelor de așteptare.

3.3.1. $M/G/1 : (\infty/GD)$

Intensitatea medie a sosirilor este din nou λ și cea a servirii μ. Vom presupune că intensitatea traficului este din nou mai mică decît unitatea. Apoi, vom considera numărul clienților existenți în sistem, imediat după o plecare a unui client din sistem.

Presupunem că un client care părăsește sistemul lasă în urma sa n alți clienți în sistem, incluzîndu-l pe cel care este servit și al cărui timp de servire va fi t; mai presupunem că sînt r noi veniți în perioada t, astfel încît numărul n' existent în momentul următoarei plecări din sistem este

$$n' = \max(n - 1, 0) + r = n + r - 1 + \delta(n), \qquad (3.15)$$

unde $\delta(n) = 0$, $n > 0$, $\delta(0) = 1$. Apare astfel evident că $[\delta(n)]^2 = \delta(n)$ și $n[1 - \delta(n)] = n$. În echilibru statistic, dacă numărul mediu de persoane care așteaptă este n, atunci $n = n'$. În mod similar $D(n) = D(n')$. Prin urmare, pe parcursul timpului de servire de durată t

$$\overline{\delta(n)} = 1 - \bar{r} = 1 - \lambda t$$

și

$$D(r) = \lambda^2 t^2 + \lambda t.$$

Deoarece timpul mediu de servire este $1/\mu$, avem

$$\bar{r} = \lambda/\mu = \rho.$$

Probabilitatea a r sosiri este independentă de n și de $\delta(n)$. Aplicînd regulile statistice obișnuite privind valorile medii ale produsului de variabile independente, avem

$$M[r^2] = \lambda^2 D(t) + \rho^2 + \rho.$$

De asemenea,

$$n'^2 = n^2 - 2n(1 - r) + (r - 1)^2 + \delta(2r - 1)$$

care, în stare de echilibru, dă

$$\mathrm{M}[n'^2] - \mathrm{M}[n^2] = 2\mathrm{M}[n]\,\mathrm{M}[r - 1] + \mathrm{M}[(r - 1)^2] +$$

$$+ \mathrm{M}[\delta(n)] \cdot \mathrm{M}[2r - 1],$$

relație care poate fi combinată cu relațiile anterioare și din care se obține

$$\bar{n} = \rho + \frac{\rho^2 - \lambda^2 \mathrm{D}(t)}{2(1 - \rho)} \tag{3.16}$$

pentru numărul mediu de persoane existente în sistem în momentele de după *plecări*. Acesta nu poate diferi cu mai mult de o persoană de numărul mediu de persoane prezentate la un moment dat. Este ușor de arătat că media timpului de așteptare este

$$W_p = \frac{\rho^2 + \lambda^2 \mathrm{D}(t)}{2\lambda(1 - \rho)}. \tag{3.17}$$

Argumentul care conduce la acest rezultat este acela că numărul mediu de sosiri pe parcursul timpului mediu de așteptare este $\lambda(W_q + 1/\mu)$, pe care îl egalăm cu $\bar{n}$. Așadar :

$$\lambda(W_q + 1/\mu) = \rho + \frac{\rho^2 + \lambda^2 \mathrm{D}(t)}{2(1 - \rho)},$$

de unde rezultă (3.17).

Aceste rezultate sînt cunoscute ca formulele Pollaczek-Hincin. Ele se utilizează pentru examinarea efectelor diferitelor repartiții ale timpilor de servire.

3.3.2. $M/E_k/1 : (\infty/\text{FIFO})$

Doar modelele cu sosiri aleatoare și repartiții exponențiale negative ale timpului de servire conduc la ecuații liniare simple. Pentru alte tipuri de repartiții, sistemul de ecuații este de obicei mult mai complicat și în general el nu poate fi rezolvat explicit cu ajutorul unor funcții simple. Majoritatea proceselor de așteptare care au loc nu au repartiții exponențiale ale timpului de servire. Ce este atunci de făcut, pentru a crea modele ușor de analizat pentru asemenea procese?

Karl Erlang, în lucrarea sa de pionierat, sugerează un răspuns la această problemă. El observă că cea mai mare parte a repartițiilor neexponențiale ale timpului de servire pe care le-a întîlnit pot fi aproximate prin funcția de repartiție cunoscută sub numele de repartiția Euler, repartiția Pearson de tipul III, repartiția χ^2, repartiția Erlang și repartiția gamma. (Ar fi interesant de aflat explicația creșterii numărului de astfel de nume.) Vom utiliza densitatea de repartiție, egală cu

$$\frac{(\mu k)^k}{(k-1)!} e^{-\mu k t} t^{k-1} \, dt.$$

Așa după cum am văzut (cap. 2) aceasta este repartiția sumei a k variabile aleatoare, fiecare repartizată exponențial cu media $1/(\mu k)$. Erlang și-a dat seama că un mecanism de servire cu o astfel de repartiție poate fi conceput ca fiind format din k etape în serie, fiecare etapă fiind o stație de servire exponențială. El nu a afirmat că mecanismul de servire necesită o structură în k faze ; tot ceea ce este necesar pentru scopuri analitice este ca acesta ar putea să se comporte *ca și cum* ar fi așa.

Cititorul va vedea că dacă un model poate fi construit pe baza elementelor exponențiale, ecuațiile ce trebuie să fie rezolvate vor fi mult simplificate.

Regulile care guvernează activitatea unei faze-tip a mecanismului de servire cu k faze (sau etape) sînt :

1. Numai un client poate fi servit la un anumit moment dat de stația de servire.
2. Fiecare client trebuie să străbată fiecare fază a serviciului, începînd cu prima.
3. După terminarea unei faze, clientul se deplasează imediat la faza următoare.

4. Atunci cînd un client părăsește faza k, următorul client poate intra în faza 1, dar nu înainte.

Cititorul va remarca că aceasta nu este același lucru cu k servicii în serie, deoarece în cazul de care ne ocupăm, numai un client poate fi servit la un anumit moment dat, în timp ce în celălalt caz puteau fi implicați în mecanismul de servire pînă la k persoane. Datorită acestei condiții date de regula 1, putem considera cele k faze echivalente cu o singură operație.

Modelul de așteptare $M/E_k/1$: (∞/FIFO) poate fi studiat utilizînd metoda Erlang. Un mod de a stabili ecuațiile constă în a nota probabilitatea ca la momentul t să existe n clienți prezenți iar clientul care este servit să se afle în faza i, cu simbolul $p_{ni}(t)$. Aceasta este totuși o complicație inutilă. Este preferabil să notăm cu $p_n(t)$ probabilitatea să existe n *faze* de așteptare și de servire la momentul t; astfel, fiecare client nou sosit va face să crească numărul de faze existent cu k, și ori de cîte ori o fază este completată, numărul lor este redus cu o unitate.

Ecuațiile stării stabile pentru această situație pot fi construite în maniera obișnuită, examinînd toate posibilitățile. Aceste ecuații sînt :

$$\lambda p_0 = k\,\mu\, p_1, \tag{3.18}$$

$$(\lambda + k\mu)p_n = k\mu p_{n+1} + \lambda p_{n-k}, \quad n \geqslant 1. \tag{3.19}$$

Funcția generatoare poate fi dedusă din aceste ecuații în mod obișnuit, prin înmulțirea membrului stîng al ecuației în p_n cu s^n și însumînd după valorile lui n, de la 0 la ∞. Astfel :

$$P(s) = \frac{p_0}{1 - \rho s \dfrac{1 - s^k}{1 - s}} = \frac{p_0(1 - s)}{1 - s - \rho s\,(1 - s)^k}. \tag{3.20}$$

Deoarece $P(1) = 1$, prin definiția funcției generatoare $p_0 = 1 - k\rho$. În aceste expresii ρ este interpretat ca fiind $\lambda/(k\mu)$, adică intensitatea traficului fiecărei faze individuale și nu este intensitatea traficului total λ/μ, corespunzător întregului mecanism de servire.

Nu este dificil, fiind doar calcule algebrice anevoioase, să determinăm expresia lui p_n din funcția ei generatoare. De fapt,

coeficientul lui s^n dă

$$p_n = (1 - k\rho) \sum \rho^m (-1)^i \binom{m}{i} \binom{m + j - 1}{j}, \tag{3.21}$$

unde însumarea trebuie făcută după toate partițiile lui n, conform relației $n = j + ik + m$. După alte calcule se găsește numărul mediu al clienților în sistem la un moment dat:

$$\bar{n} = \frac{1}{2} \frac{(\lambda/\mu)}{1 - (\lambda/\mu)} \frac{k+1}{k}. \tag{3.22}$$

Repartiția Erlang sau Pearson de tipul III, care reprezintă repartiția timpului total de servire în cele k faze, are evident valoarea medie $1/\mu$ și dispersia $(1/k)(1/\mu^2)$. Prin urmare, dacă coeficientul de variație este notat cu V_k, atunci $V_k^2 = 1/k$. Așadar, vom avea

$$\frac{k+1}{k} = 1 + V_k^2$$

și formula pentru numărul mediu de clienți din sistem poate fi scrisă astfel:

$$\bar{n} = \frac{1}{2}(1 + V_k^2) \frac{(\lambda/\mu)}{1 - (\lambda/\mu)},$$

care este exact rezultatul dat de formula Pollaczek-Hincin. Este o verificare utilă a rezultatului, mai cu seamă că interpretările legate de semnificația lui p_n sînt diferite în cele două situații. În deducerea formulei Pollaczek-Hincin, p_n a fost ales ca fiind probabilitatea ca la plecarea unui client, acesta să lase pe altcineva în urma sa. În situația noastră, p_n a fost considerat a fi proporția timpului pentru care vor fi prezenți n clienți în sistem.

Timpul mediu de așteptare este dat de

$$W_q = \frac{1}{2\mu}(1 + V_k^2) \frac{(\lambda/\mu)}{1 - (\lambda/\mu)}; \tag{3.23}$$

Este limpede că în condiţiile în care k tinde la infinit, V_k va tinde către zero, iar formulele vor da rezultate pentru timpii de servire *constanţi.*

Repartiţia timpului de aşteptare poate fi obţinută considerînd că n reprezintă numărul fazelor existente în sistem şi că fiecare fază este exponenţială. Timpul total de aşteptare pentru ultimul client, considerînd ca disciplină de aşteptare regula FIFO, este suma a n variabile exponenţiale, fiecare avînd media $1/\mu$. Deci transformata Laplace a repartiţiei timpului de aşteptare este dată de relaţia

$$W_q^*(z) = \sum_{n=0}^{\infty} p_n \left(\frac{1}{1+z/\mu}\right)^n.$$

Reamintind că $P(s)$ este funcţia generatoare a lui p_n, aceasta este echivalentă cu

$$P\left(\frac{1}{1+z/\mu}\right). \tag{3.24}$$

Momentele repartiţiei timpului de aşteptare pot fi calculate din această expresie, prin derivare repetată, în raport cu z şi făcînd $z = 1$ în rezultate. Această inversare a transformatei Laplace ar fi de un risc enorm, chiar folosind tehnicile aproximative. De aceea, repartiţia timpului de aşteptare nu este uşor de obţinut, exceptînd cazurile în care numărul fazelor este relativ mic (să spunem, $k = 2$ sau 3) *.

3.4. MODELE CU TIMPUL DE SERVIRE CONSTANT

3.4.1. $M/D/1 : (\infty/\text{FIFO})$

În capitolul următor ne vom ocupa sumar de analizele făcute de Crommelin asupra modelului $M/D/c : (\infty/\text{FIFO})$; fiind un caz special al acestui model, anume pentru $c = 1$, nu vom insista în a demonstra rezultatele ce vor urma.

* Vezi Morse, pp. 72–82.

Funcția generatoare pentru acest model cu o stație de servire este

$$P(s) = \frac{(1-\rho)(1-s)}{1 - s\,e^{\rho(1-s)}}. \tag{3.25}$$

Prin derivarea de n ori în raport cu s, împărțind apoi prin n și făcînd $s = 0$, obținem

$$p_0 = 1 - \rho; \quad p_1 = (1-\rho)(e^{\rho} - 1);$$
$$p_n = (1-\rho)\sum_{k=1}^{n}(-1)^{n-k}e^{k\rho}\left[\frac{(k\rho)^{n-k}}{(n-k)!} + \frac{(k\rho)^{n-k-1}}{(n-k-1)!}\right] \quad \text{pentru } n \geqslant 2. \tag{3.26}$$

Aceasta este forma generală a soluției stării stabile. Prin derivarea în raport cu s a funcției generatoare, punînd apoi $s = 1$, obținem că numărul mediu de clienți în sistem este

$$\bar{n} = 1 + \frac{1}{2}\,\frac{\rho}{1-\rho}.$$

Deci timpul mediu de așteptare în șir și timpul mediu total de așteptare pot fi deduse ușor:

$$W_q = \frac{1}{2\mu}\,\frac{\rho}{1-\rho}, \qquad W = \frac{1}{2\mu}\,\frac{1}{1-\rho}.$$

Repartiția timpului de așteptare poate fi obținută și în forma

$$W_q(v) = (1-\rho)\sum_{i=0}^{k} e^{\rho(\mu v - i)}\,\frac{[-\rho(\mu v - i)]^i}{i!}. \tag{3.27}$$

Acest model este deosebit de util, deoarece există numeroase situații reale în care coeficientul de variație a repartiției timpului de servire este destul de mic, astfel încît timpul de servire poate fi considerat ca fiind constant. În cadrul acestei cărți vom avea ocazia să studiem cel puțin un astfel de caz.

3.4.2. $M/D/1 : (\infty/\text{SIRO})$

După cum arătam, în cadrul multor fenomene cotidiene, regula FIFO a disciplinei de aşteptare nu este respectată; de exemplu în încercarea de a lua maşina în timpul orelor de vîrf, de a cumpăra alimente dintr-un magazin etc. În astfel de circumstanţe, ordinea de servire degenerează în aleatoare şi devine posibilă aplicarea regulii SIRO. De aceea, modelul prezent care combină timpii constanţi de servire cu servirea în ordine aleatoare devine uneori foarte folositor.

Dacă nu se cunoaşte cu exactitate disciplina de aşteptare, este indicat a prelucra rezultatele atît de la modelul $M/D/1 :(\infty/\text{FIFO})$ cît şi de la $M/D/1 :(\infty/\text{SIRO})$ şi de a le considera ca fiind limitele extreme ale unui rezultat corect. Acestea au fost consideraţiile care l-au condus pe Burke * să cerceteze acest model. Lucrarea sa nu prezintă numai o analiză teoretică ci, de asemenea, un util set de grafice care indică :

1. Probabilităţile $|1 - W_e(v/b)|$ pentru valori mai mici decît $130b$.

2. Comparaţia timpilor de aşteptare pentru cele două modele $M/D/1 : (\infty/\text{FIFO})$ şi $M/D/1 :(\infty/\text{SIRO})$.

3. Comparaţia timpilor de aşteptare pentru cele două modele $(M/M/1 : (\infty/\text{SIRO})$ şi $M/D/1 : (\infty/\text{SIRO})$.

Înainte de a prezenta analiza făcută de Burke, vom rezuma concluziile acestor grafice. În primul rînd, comparaţia 2 indică, aşa cum ne aşteptăm, că regula SIRO determină o repartiţie a timpilor de aşteptare de alt fel decît regula FIFO. Există mai mulţi timpi scurţi şi lungi şi mai puţini în mijlocul domeniului repartiţiei (ceea ce este fără îndoială adevărat pentru aceste reguli). Această deosebire merită luată în considerare. Ea este foarte mică pentru durate scurte, adică pentru timpi de aşteptare pînă în punctul în care curbele celor două modele se intersectează. Dar deosebirea creşte pronunţat pe măsura creşterii timpilor de aşteptare. De exemplu, la o intensitate a traficului de 0,7, probabilitatea unui timp de aşteptare egal cu 10 timpi medii de servire este de zero ori mai mare pentru modelul SIRO, faţă de FIFO. La o intensitate a traficului de 0,8, probabilitatea unei

---

* Burke, P. J. *Equilibrium Delay Distribution for One Channel with constant Helding Time Poisson Input and Random Service.* În : Bell System Tech. J., July, 1959. Graficele din lucrarea lui Burke sînt reproduse în această carte în anexa 2, graficele A2-1—A2-4.

întîrzieri similare este mai mică de trei ori pentru SIRO față de FIFO. Pentru $\rho = 0{,}9$, probabilitățile sînt aproape identice.

Comparația 3 făcută de Burke este extrem de interesantă, deoarece ea conduce la o foarte utilă regulă de lucru. Pentru orice nivel dat al probabilității, timpul corespunzător de așteptare pentru un timp constant de servire este aproximativ egal cu jumătatea celui determinat pentru modelul cu timp exponențial de servire. Cîteva exemple sînt date în tabelul 3.1.

Tabelul 3.1

Comparație între $M/M/1$: (∞/SIRO) și $M/D/1$: (∞/SIRO)

Nivelul probabilității	ρ	v pentru timpi constanți de servire	v pentru timpi exponențiali de servire
0,1	0,9	12μ	24μ
0,1	0,6	$2{,}5\mu$	5μ
0,01	0,9	37μ	78μ
0,01	0,6	7μ	14μ
0,001	0,8	32μ	66μ
0,001	0,6	12μ	27μ

Deducerea repartiției timpilor de așteptare făcută de Burke începe cu ecuația Crommelin pentru $c = 1$:

$$p_{n-1} = (p_0 + p_1) \cdot \frac{\lambda^{n-1} e^{-\lambda}}{(n-1)!} + \ldots + p_n e^{-\lambda}.$$

El introduce apoi probabilitatea condiționată $q(t|n)$, ca timpul de așteptare să nu fie mai mare decît t în cazul în care clientul care așteaptă este unul din cei n, la momentul următor unei plecări. Fie $F(t)$ repartiția timpilor de așteptare. Atunci

$$F(t) = \sum_{n=1}^{\infty} p_{n-1}\, q(t|n)$$

și problema este de a evalua $q(t|n)$.

Intervalele de timp dintre o sosire și prima plecare din sistem trebuie să fie repartizate uniform deoarece intrările sînt repartizate

Poisson și trebuie să fie independente de intervalele imediat următoare, deoarece disciplina de așteptare este SIRO. În acest caz, timpul total de așteptare poate fi exprimat ca o sumă de două elemente

$$v = v_1 + v_2,$$

unde v_1 este multiplu întreg al timpului mediu de servire luat ca unitate de timp și v_2 este partea fracționară rămasă. În mod analog, timpul t poate fi divizat în t_1 și t_2. Utilizînd independențele subliniate mai înainte, avem

$$\begin{aligned} q(t\,|\,n) &= P(v \leqslant t\,|\,n) \\ &= P(v_1 < t|n) + P(v_1 = t_1|n)\, P(v_2 = t_2). \end{aligned}$$

și, ținînd cont că v_2 este repartizat uniform,

$$q(v_1\,|\,n) = \sum_{i=0}^{n-1} P(v_1 = i\,|\,n) + v_2\, P(v_1 = t_1\,|\,n).$$

Notînd $P(v_1 = i|n) = Q_i(n)$, obținem $Q_0(n) = 1\,|\,n$ și

$$Q_i(n) = [1 - Q_0(n)] \sum_{j=0}^{\infty} \frac{\lambda j e^{-\lambda}}{j!} Q_{i-1}(n + j - 1)\ (i = 1, 2, \ldots). \quad (3.28)$$

Această formulă de recurență ne va da $q(v|n)$ și eventual pe $W_q(v)$, dacă p_n a fost calculat anterior. În acest mod a obținut Burke graficele la care ne-am referit.

3.5. REGULI DE PRIORITĂȚI PENTRU DISCIPLINA DE AȘTEPTARE

Analizele făcute de Burke, legate de efectele regulii SIRO în cazul modelelor cu un singur șir de așteptare și cu timp constant de servire reprezintă un moment prielnic pentru a discuta regulile de

prioritate pentru disciplina de așteptare, cu aceasta încheind capitolul de față.

În prezenta carte nu ne vom ocupa în mod detaliat de priorități, dar aceasta nu înseamnă că subiectul este de neglijat. În liniile tehnologice ca și în activitățile legate de întreținere și reparația utilajelor, stabilirea priorităților anumitor componente reprezintă principala cale pentru a asigura eficiența producției și a minimiza cheltuielile de întreținere. Aceste aspecte însă nu vor fi discutate în cadrul acestei cărți. Din această cauză, actuala prezentare va fi redusă la minimum.

3.5.1. PRIORITĂȚI ABSOLUTE

Să presupunem două clase de clienți C_1 și C_2. Pentru toți clienții care intră în sistemul de așteptare, α aparțin lui C_1 și restul de $(1-\alpha)$ lui C_2. Este necesar să considerăm doar două clase de priorități, deoarece este foarte ușor să generalizăm rezultatele obținute pentru două clase și în cazul priorităților pentru mai multe clase. Cînd un client din C_1 sosește, el are prioritate față de toți membrii clasei C_2 care există în sistem, dar se găsește în urma celorlalți membri ai clasei C_1 existenți deja în sistem.

Notăm cu $p_i(n, m)$ probabilitatea, în starea stabilă, ca un client aparținînd clasei i să fie servit cînd sînt n clienți din clasa C_1 în sistem și m clienți din clasa C_2 în sistem. Atunci, ecuațiile stării stabile vor fi scrise după cum urmează, dacă intensitatea medie de servire este aceeași pentru ambele clase:

$$p_1(1,0) + p_2(0,1) = \rho\, p_0(0,0) = \rho K_0,$$

$$\alpha\rho K_0 + p_1(2,0) + p_2(1,1) = (\rho+1)p_1(1,0),$$

$$(1-\alpha)\rho K_0 + p_1(1,0) + p_2(0,2) = (\rho+1)p_2(0,1),$$

$$\alpha\rho p_1(n-1,0) + p_1(n+1,0) + p_2(n,1) = (\rho+1)p_1(n,0),$$

$$(1-\alpha)\rho p_2(0,n-1) + p_1(1,m) + p_2(0,m+1) = (\rho+1)\,p_2(0,m),$$

$$(1-\alpha)\rho p_1(1,m-1) + p_1(2,m) + p_2(1,m+1) = (\rho+1)p_1(1,m),$$

$$\alpha\rho p_2(n-1,1) = (\rho+1)\,p_2(n,1),$$

$$\alpha\rho p_1(n-1, m) + (1-\alpha)\rho p_1(n, m-1) + p_1(n+1, m) + p_2(n, m+1) = (\rho+1)\, p_1(n, m),$$
$$(n>1;\ m>0),$$
$$\alpha\rho p_2(n-1, m) + (1-\alpha)\rho p_2(n, m-1) = (\rho+1)p_2(n, m),$$
$$(n>0;\ m>1). \tag{3.29}$$

Aceste ecuații pot fi rezolvate prin metoda cunoscută a determinării funcției generatoare, care în acest caz este în două variabile: în ξ alocat pentru n și s pentru m. Subliniem că, în mod necesar, deoarece intensitățile medii de servire sînt identice pentru C_1 și C_2, avem $K_0 = 1 - \rho$. Mai mult,

$$\sum_{r=0}^{n-1} \{p_1(n-r, r) + p_2(r, n-r)\} = (1-\rho)\rho^n.$$

Calculele algebrice legate de determinarea funcțiilor generatoare sînt foarte anevoioase și nu vor fi reproduse aici. Din funcția generatoare se obțin următoarele rezultate:

$$\bar{n} = \alpha\rho\,\frac{1+\rho-\alpha\rho}{1-\alpha\rho}, \tag{3.30}$$

$$\bar{m} = (1-\alpha)\,\rho\,\frac{1-\alpha\rho+\alpha\rho^2}{(1-\rho)(1-\alpha\rho)}, \tag{3.31}$$

$$\bar{n}_q = \frac{\alpha\rho^2}{1-\alpha\rho}, \tag{3.32}$$

$$\bar{m}_q = \frac{(1-\alpha)\rho^2}{(1-\rho)(1-\alpha\rho)}, \tag{3.33}$$

$$W_{q1} = \frac{\lambda}{\mu(\mu-\alpha\lambda)}, \tag{3.34}$$

$$W_{q2} = \frac{\lambda}{(\mu-\lambda)(\mu-\alpha\lambda)}. \tag{3.35}$$

Este interesant de remarcat că raportul timpilor medii de așteptare pentru cele două clase de clienți este

$$\frac{\mu}{\mu - \lambda},$$

care este independent de α. Nu este surprinzător să constatăm că atunci cînd intensitatea traficului se apropie de 1, șirul clasei C_2 va deveni extrem de mare, în timp ce pentru clienții clasei C_1 tinde spre limita stării stabile.

Cox și Smith analizează în detaliu situația în care există k clase de priorități. Ei arată că în această situație:

$$W_{qj} = \frac{\frac{1}{2}\sum_{i=1}^{k} \alpha_i c_i}{\left(1 - \sum_{i=1}^{j-1} \frac{\alpha_i}{\mu_i}\right)\left(1 - \sum_{i=1}^{j} \frac{\alpha_i}{\mu_i}\right)}, \tag{3.36}$$

unde α_i este proporția clienților care aparțin clasei C_i, μ_i este intensitatea medie de servire a acestor clienți, iar c_i este momentul de ordin doi al repartiției timpului de servire pentru membrii clasei C_i. De asemenea, ei arată că timpul mediu de așteptare pentru toți clienții este $W_q = \sum_{i=1}^{k} \alpha_j W_{qj}$.

Aceste expresii pot fi folosite în evaluarea sistemelor optime de alocare a priorităților. În mod special, dacă întîrzierilor în șir le pot fi repartizate penalități astfel încît valorile lor să fie proporționale cu timpii medii de așteptare ai diferitelor clase, atunci prioritățile ar putea fi repartizate în concordanță cu valorile rapoartelor:

$$\frac{\text{timpul mediu de servire}}{\text{penalitatea pe unitate de timp de așteptare}}.$$

Cu cît valoarea raportului este mai mică, cu atît va fi mai mare prioritatea.

Aceasta este tot ce am vrut să prezentăm legat de prioritățile absolute. Cititorul care dorește să studieze problema mai profund este sfătuit să consulte lucrările citate în cadrul bibliografiei selectate la sfîrșitul acestei cărți. Recomandăm cu căldură explicațiile date în cartea lui Cox și Smith.

3.5.2. PRIORITĂȚI RELATIVE

Această secțiune va începe cu un comentariu personal. Exceptînd electronica și telecomunicațiile, nu am întîlnit situații în care să fi fost utilizate prioritățile relative sau să fi existat posibilitatea utilizării lor, chiar atunci cînd acest lucru ar fi fost de dorit. Nu cred că prioritățile relative să aibă aplicații în procese de așteptare în care clienții sînt ființe umane. Deoarece personal am avut puține ocazii în care am întîlnit priorități relative, s-ar putea să fie foarte dificil pentru cititor să-și imagineze situații în care pot să apară. O aplicație majoră este legată de utilizarea sistemelor mari de calculatoare electronice digitale care operează on-line. Pentru sistemele on-line nu este deloc neobișnuit să aibă de rezolvat o mare varietate de probleme. Unele dintre ele sînt mai importante decît altele și sînt repartizate priorități relative, astfel încît dacă o problemă prioritară necesită o prelucrare imediată, o rutină specializată a sistemului de operare memorează programul care era rulat în acel moment, întrerupîndu-l, și-l încarcă pentru execuție pe cel cu prioritate mai mare. Această situație este foarte utilizată la sistemele electronice on-line, destinate rezervărilor de bilete pentru cursele aeriene. În astfel de cazuri, după ce o problemă cu prioritate înaltă a fost prelucrată, programul întrerupt revine ca și cum acum prioritatea sa ar fi cea mai înaltă, dar imediat prioritatea sa reintră în nivelul normal.

În cadrul sistemelor cu prioritate relativă clienții cu maximă prioritate nu sînt afectați deloc de cei cu priorități mai scăzute. Să presupunem de exemplu că, asemănător secțiunii precedente, există două clase prioritare C_1 și C_2, în proporțiile α și $1 - \alpha$ ale clienților. Timpii lor de servire sînt repartizați exponențial, cu parametrii μ_1 și μ_2. Atunci, probabilitatea, în stare stabilă, ca să existe n clienți din C_1 în sistem este

$$p_{1n} = \rho_1^n(1 - \rho_1), \quad n = 0, 1, 2, \ldots,$$

unde $\rho_1 = \alpha\lambda/\mu_1$. Ecuațiile stării stabile pot fi scrise ca de obicei pentru probabilitățile $p(n, m)$ ca să existe n clienți ai clasei prioritare C_1 și m din C_2 în sistem. Funcția generatoare poate fi dedusă destul de ușor, iar apoi prin derivare poate fi arătat că numărul mediu al clienților din C_2 este

$$\bar{m} = \frac{\rho_2}{1 - \rho_1 - \rho_2}\left(1 + \frac{\rho_1}{1 - \rho_1} \cdot \frac{\mu_2}{\mu_1}\right).$$

De asemenea, este posibil a deduce timpii medii de așteptare și timpii medii totali de așteptare în felul următor:

$$W_1 = 1 + \frac{1}{1-\rho}\frac{1}{\mu_1}; \qquad W_2 = \frac{1}{\mu_2(1-\rho_1-\rho_2)}\left(1 + \frac{\rho_1}{1-\rho_1}\frac{\mu_2}{\mu_1}\right);$$

$$W_{q1} = \frac{\rho_1}{\mu_1(1-\rho_1)}; \qquad W_{q2} = \frac{1}{\mu_2(1-\rho_1-\rho_2)}\left(\rho_1+\rho_2+\frac{\rho_1}{1-\rho_1}\frac{\mu_2}{\mu_1}\right).$$

3.6. COMENTARIU GENERAL

Există un model foarte general al proceselor cu un șir de așteptare, cu intrări și servire generale și disciplina de așteptare FIFO, elaborat de W. L. Smith. Este posibil — cu unele eforturi însă — a deduce rezultatele FIFO determinate în cadrul acestui capitol din formulele generale ale lui Smith. Nu există însă asemenea analize pentru modelele cu mai multe șiruri de așteptare. Cititorul, care este familiarizat cu transformata Laplace și teoria funcțiilor de variabilă complexă, este călduros sfătuit să citească lucrarea lui Smith, în cazul fericit că ar putea-o procura.

4

MODELE CU MAI MULTE STAȚII DE SERVIRE

4.1. MODELE

Diferența esențială dintre sistemele cu o stație de servire și cele cu mai multe stații de servire este aceea că ultimele prezintă o mai mare varietate a disciplinelor de așteptare; trebuie ținut seama mai mult de posibilitatea unei comportări neobișnuite a unui client.

Nu este deloc ușor de a formula reguli matematice în cazul modelelor cu tipurile cele mai amorfe de disciplină a așteptării. Pînă în prezent s-au făcut foarte puține încercări în acest sens. Modelele clasice cu mai multe stații de servire presupun de obicei regula FIFO și mai rar regulile LIFO sau SIRO. În orice caz s-a presupus principial că toți clienții sosiți vor forma un singur șir de așteptare și că acest șir va alimenta cu clienți stațiile de servire, după o anumită regulă bine definită de tipul FIFO, LIFO sau SIRO. Posibilitatea dezintegrării — unicul șir să se împartă în mai multe șiruri, poate chiar cîte unul pentru fiecare stație de servire — precum și fenomenele asociate, ca cel de deplasare, servire preferențială etc., nu au fost luate în considerare. Chiar și în aceste situații, modelele cu mai multe stații de servire sînt mult mai greu de analizat din punct de vedere matematic decît cele cu o singură stație. Din această cauză au fost cercetate foarte puțin și dispunem de foarte puține modele utilizabile.

4.1.1. $M/M/c$: (∞/FIFO)

Acesta este cel mai simplu model cu mai multe stații de servire și de aceea îl vom prezenta la început. Se presupune că sosirea clien-

ților are loc după o repartiție Poisson, cu o intensitate medie de λ pe minut. La sosire, fiecare client are trei opțiuni de comportament :

1. Dacă toate stațiile de servire sînt ocupate, se va duce la unicul șir de la care se alimentează toate stațiile.

2. Dacă o stație este liberă și toate celelalte ocupate, se va duce imediat la cea liberă.

3. Dacă există două sau mai multe stații libere, se va duce la oricare din ele, fără vreo preferință.

Pe măsură ce stațiile se eliberează, clientul care a stat cel mai mult în așteptare se duce de îndată la stația eliberată ; dacă nu există clienți în așteptare, evident că stația va rămîne liberă pînă la o nouă sosire, iar atunci, conform cu regulile descrise, clientul nou sosit va putea alege serviciul acesteia. Nu există posibilitatea ca o stație să fie liberă dacă sînt clienți în așteptare.

Toate stațiile realizează același fel de servicii. În cadrul fiecărei stații, repartiția timpului de servire este exponențială cu intensitatea b, astfel încît intensitatea medie a serviciului este de μ clienți pe minut.

Ca de obicei, vom nota cu $p_n(t)$ probabilitatea prezenței a n clienți în sistem la momentul t. Cu această notație este ușor să deducem ecuațiile diferențiale pentru modelul pe care-l analizăm, urmînd în linii mari raționamentul utilizat în capitolul precedent pentru modelul $M/M/1:(\infty/\text{FIFO})$. Acestea sînt :

$$\frac{dp(0,t)}{dt} = -\lambda p(0,t) + \mu p(1,t),$$

$$\frac{dp(n,t)}{dt} = -(\lambda + n\mu)\,p(n,t) + \lambda p(n-1,t) + (n+1)\,\mu p(n+1,t);\quad 1 \leqslant n \leqslant c-1, \tag{4.1}$$

$$\frac{dp(n,t)}{dt} = -(\lambda + c\mu)\,p(n,t) + \lambda p(n-1,t) + c\mu p(n+1,t),\quad c \leqslant n.$$

Soluțiile dependente de timp ale acestor ecuații pot fi obținute cu ajutorul transformatei Laplace a funcției generatoare — notată cu $P^*(s,z)$ — și, ca atare, pot fi deduse momentele repartiției numărului

de clienţi existenţi în sistem la momentul t. Aceste momente pot fi folosite pentru obţinerea unei repartiţii aproximative. Totuşi, nu este uşor să determinăm aceste soluţii dependente de timp şi, după cum am văzut, aceste soluţii nu prezintă un interes practic deosebit. Prin urmare, vom trece încă o dată la determinarea soluţiei stării stabile, prin eliminarea lui t şi egalînd derivatele cu zero. Va rezulta că :

$$
\begin{aligned}
&-\lambda p_0 + \mu p_1 = 0\,;\\
&-(\lambda + n\mu)\,p_n + \lambda p_{n-1} + (n+1)\,\mu p_{n+1} = 0,\ 1 \leqslant n \leqslant c-1\,; \qquad (4.2)\\
&-(\lambda + c\mu)\,p_n + \lambda p_{n-1} + c\mu p_{n+1} = 0,\quad c \leqslant n.
\end{aligned}
$$

Se poate demonstra, prin inducţie, că probabilităţile stării stabile date de aceste ecuaţii sînt :

$$
\begin{aligned}
p_0 &= \frac{1}{\sum_{r=0}^{c-1} \frac{(c\rho)^r}{r!} + \frac{(c\rho)^c}{c!}\,\frac{1}{1-\rho}}\,;\\
p_i &= \frac{1}{i!}(c\rho)^i\,p_0, \quad i \leqslant c\,; \qquad (4.3)\\
p_i &= \frac{c^c}{c!}\,\rho^i\,p_0, \qquad i > c.
\end{aligned}
$$

În aceste formule $\rho = \lambda/(c\mu)$, adică traficul măsurat în erlangi împărţit cu numărul staţiilor de servire. Traficul obţinut este uneori notat cu ε, iar formulele de mai sus sînt atunci scrise înlocuindu-l pe ρ cu ε/c. (Excelenta carte a lui Fry utilizează această notaţie.)

Acest model presupune că nu există limitare a clienţilor care pot exista în sistem la orice moment dat ; altfel spus, sistemul are *o capacitate infinită*. Aceasta este o simplificare ce poate fi înlăturată cu uşurinţă, iar modelul rezultat este mai general şi mai folositor.

Să presupunem că prima din cele trei reguli legate de comportamentul clientului este înlocuită de următoarea :

1a. Dacă la sosirea unui client toate stațiile de servire sînt ocupate, acesta se va așeza la unicul șir de așteptare al tuturor stațiilor dacă numărul clienților existenți în sistem este mai mic decît L, iar dacă acest număr este L, el nu se va mai așeza în șir și va părăsi sistemul.

Vom numi aceasta, *restricția asupra capacității sălii de așteptare.* Capacitatea maximă a sistemului este L, iar capacitatea maximă a sălii de așteptare este $L - c$. Clienții care sosesc și, găsind sistemul supraaglomerat, îl părăsesc, sînt numiți clienții respinși sau, în cazul sistemului telefonic, *convorbiri pierdute.*

Pentru acest model, ecuațiile stării stabile sînt :

$$\begin{aligned}
-(\lambda + n\mu)p_n + \lambda p_{n-1} + (n+1)\,\mu p_{n+1} &= 0, \quad (1 \leqslant n \leqslant c-1) \\
-(\lambda + c\,\mu)\,p_n + \lambda p_{n-1} + c\mu p_{n+1} &= 0, \quad (c \leqslant n \leqslant L-1) \\
-c\mu p_L + \lambda p_{L-1} &= 0, \\
-\lambda p_0 + \mu p_1 &= 0.
\end{aligned} \tag{4.4}$$

Din nou aceste ecuații pot fi ușor rezolvate, rezultînd următoarele probabilități :

$$p_0 = \frac{1}{\sum_{r=0}^{c-1} \frac{(c\rho)^r}{r!} + \frac{(c\rho)^c}{c!} \cdot \frac{1-\rho^{L-c+1}}{1-\rho}};$$

$$p_i = \frac{1}{i!}\,(c\rho)^i\,p_0, \quad i = 1, 2, \ldots, c; \tag{4.5}$$

$$p_j = \frac{c^c}{c!}\,\rho^j\,p_0, \quad j = c+1, \ldots, L.$$

Trăsătura interesantă a acestui model este că, pentru orice trafic obținut λ/μ, există două posibilități de manevrare a probabilităților de așteptare : prin ajustarea capacității sistemului sau a numărului de stații de servire. Cititorul va remarca că atunci cînd capacitatea sis-

temului poate fi făcută oricît de mare, se obţin încă o dată formulele modelului Erlang $M/M/c:(\infty/\text{FIFO})$. Iar dacă luăm $c = 1$, vom obţine

$$p_0 = \frac{1 - \rho}{1 - \rho^{L+1}},$$

$$p_i = \rho^i p_0, \qquad i = 1, 2, \ldots, L,$$

care reprezintă tocmai probabilităţile stării stabile pentru modelul cu o staţie de servire avînd capacitate limitată. Făcîndu-l pe L să tindă spre infinit, obţinem probabilităţile pentru modelul $M/M/1:(\infty/\text{GD})$. Acestea pot fi comparate cu rezultatele obţinute în capitolul precedent.

Pînă acum nu a fost necesar să utilizăm regula FIFO: „cînd o staţie se eliberează, clientul care a stat cel mai mult în aşteptare se duce de îndată la staţia eliberată". Probabilităţile pe care le-am dedus nu sînt afectate de disciplina de aşteptare şi pot fi aplicate la toate cele trei modele principale $M/M/c:(L/\text{FIFO})$, $M/M/c:(\text{L/SIRO})$ şi $M/M/c:(L/\text{LIFO})$. De aceea este limpede că numărul mediu al clienţilor existenţi în sistem nu va fi afectat de disciplina de aşteptare şi atît timp cît repartiţia timpului de servire este exponenţială, cum este în acest caz (aceste consideraţii vor trebui revăzute), timpul mediu total de aşteptare şi timpul mediu de aşteptare de asemenea trebuie să fie independente de disciplina de aşteptare. Dacă este aşa, atunci remarcăm din nou că pentru modelele cu mai multe staţii de servire efectul diferitelor discipline de aşteptare este doar alterarea probabilităţilor timpilor scurţi şi lungi de aşteptare.

Atunci cînd se aplică regula FIFO, repartiţia timpului de servire poate fi obţinută după cum urmează. Probabilitatea ca un client nou sosit să găsească sistemul supraaglomerat şi să fie respins este

$$p_L = \frac{\rho^L (c^c/c!)}{\sum_{r=0}^{c-1} \frac{(c\rho)^r}{r!} + \frac{(c\rho)^c}{c!} \frac{1 - \rho^{L-c+1}}{1 - \rho}}.$$

Probabilitatea ca să găsească cel puțin o stație de servire liberă și deci ca timpul său de așteptare să fie zero este

$$\sum_{r=0}^{c-1} p_r = \frac{\sum_{r=0}^{c-1} \frac{(c\rho)^r}{c!}}{\sum_{r=0}^{c-1} \frac{(c\rho)^r}{r!} + \frac{(c\rho)^c}{c!} \frac{1-\rho^{L-c+1}}{1-\rho}}.$$

Probabilitatea ca să găsească toate stațiile ocupate și r clienți așteptînd înaintea sa este

$$p_{c+r} = \frac{c^c}{c!} \rho^{c+r} \frac{1}{\sum_{j=0}^{c-1} \frac{(c\rho)^j}{j!} + \frac{(c\rho)^c}{c!} \frac{1-\rho^{L-c+1}}{1-\rho}}.$$

Ieșirea clienților din stațiile ocupate este repartizată Poisson, astfel încît probabilitatea ca j clienți să fie serviți într-o perioadă t este

$$e^{-c\mu t} \frac{(c\mu t)^j}{j!}.$$

Prin urmare, dacă un client nou sosit găsește r alți clienți în așteptare înaintea sa, probabilitatea ca el să stea în așteptare mai mult de t momente înainte să intre în stația de servire este dată de

$$\sum_{j=0}^{r} e^{-c\mu t} \frac{(c\mu t)^j}{j!}.$$

Înmulțind cu p_{c+r} pentru a face relația necondițională și apoi însumînd după r, între 0 și $L - c - 1$, obținem probabilitatea ca un client care intră în sistem, dar care are de așteptat, să stea în așteptare și după trecerea unei durate de timp t. (Cititorul va observa că suma trebuie să se oprească la valoarea $r = L - c - 1$, în caz contrar clientul va fi respins.) Utilizînd notațiile uzuale, obținem

$$1 - W_q(t) = P(>t) = \frac{(c\rho)^c}{c!} \frac{e^{-c\mu t}}{1-\rho} \left[\sum_{r=0}^{L-c-1} \frac{(\lambda t)^r}{r!} - \rho^{L-c} \sum_{r=0}^{L-c-1} \frac{(c\mu t)^r}{r!} \right] p_0. \tag{4.6}$$

Aceste expresii sînt destul de sofisticate, dar pot fi ușor evaluate în cazuri particulare. În felul acesta, putem calcula două măsuri ale

efectivității — sau standarde de serviciu — pentru acest model și anume :

1. Probabilitatea ca un client să stea în așteptare mai mult decît un timp specificat, să spunem τ minute, sau $\tau\mu$ timpi medii de servire.

2. Probabilitatea pierderii — sau respingerii — care este probabilitatea ca un client să găsească sistemul supraaglomerat.

Capacitatea sistemului și numărul stațiilor pot fi manevrate în scopul modificării standardelor acestor două categorii, pentru orice trafic dat. De exemplu, probabilitatea respingerii poate fi redusă prin mărirea capacității L a sistemului, dar în aceste condiții timpul mediu de așteptare și probabilitatea existenței unor timpi de așteptare de lungă durată va crește. Micșorîndu-l pe L, timpul mediu de așteptare poate fi redus, dar atunci va crește riscul respingerii. O creștere a numărului de stații, în caz că λ, μ și L sînt fixați, va reduce probabilitatea timpilor de lungă durată, cît și pe cea a respingerii. Într-o situație concretă s-ar putea să fie mai economic să se mărească capacitatea sălii de așteptare decît să se introducă stații suplimentare. De aceea, este important a decide optimul balanței între numărul de stații de servire, pe de o parte, și capacitatea sistemului, pe de altă parte.

Acest model este destul de util în practicile comerciale și de aceea este necesar a calcula pentru diferite valori ale lui λ, μ, L și c probabilitățile de felul de mai sus, care măsoară standardele de servire. Din păcate însă, este dificil a se realiza grafice utile, datorită numărului mare de factori ce trebuie luați în considerare. (La Compania de transporturi aeriene Air Canada am realizat un sistem de programe pentru calculator, menit a realiza toate aceste calcule. Acesta este un mijloc pe care-l recomandăm cititorului.)

Cînd disciplina de așteptare este de tip SIRO, nu sînt necesare calcule speciale pentru a obține formulele din cazul modelului cu mai multe stații de servire. Morse a demonstrat că formula lui Riordan pe care am întîlnit-o în capitolul anterior legat de modelul $M/M/1$: (∞/SIRO) este aplicabilă și modelului $M/M/c$: (∞/SIRO), făcînd doar modificări minore. Această formulă este

$$P(>t) = 1 - W_q(t) = \frac{1}{2}(1-\sqrt{\rho/2})\exp\{-t(1-\rho)(1-\sqrt{\rho/2}\} + \\ + \frac{1}{2}(1+\sqrt{\rho/2})\exp\{-t(1-\rho)(1+\sqrt{\rho/2}\}. \tag{4.7}$$

unde considerăm pe ρ ca fiind $\lambda/(c\mu)$. În scopul calculării lui $p_0, W_q(t)$ și W_q pentru astfel de modele, au fost realizate tabele remarcabile, datorate lui R. Wilkinson. Din păcate, nu există asemenea tabele sau grafice pentru $M/M/c$: (L/SIRO). Se pare că nimeni nu a obținut formulele pentru acest model.

Repartiția timpilor de așteptare în cazul unei reguli LIFO este aceeași cu repartiția perioadei ocupate atunci cînd se aplică regula FIFO. Aceasta nu este însă de mare folos, atît timp cît nimeni nu a putut însă obține o formulă explicită a repartiției perioadei ocupate pentru modelul $M/M/c$: (L/FIFO) chiar și pentru cazul $L = \infty$. Calculul depinde de soluțiile dependente de timp ale modelului și acestea se pot obține numai sub forma transformatelor Laplace ale funcțiilor generatoare sau a transformatelor Laplace ale probabilităților dependente de timp. Formulele ce rezultă sînt foarte sofisticate și dificil de evaluat. În anexa 1 sînt prezentate unele considerații privind tratarea matematică a acestor probleme; în practică, prin intermediul simulării, este mult mai ușor și mai rapid de a studia cazurile particulare, acest lucru avînd și avantajul că limitările privitoare la capacitatea sistemului pot fi modificate fără dificultate. În cap. 6 ne vom ocupa mai pe larg de unele aspecte legate de simulare.

4.1.2. $M/E_k/c$: (∞/FIFO)

În principiu nu prezintă dificultăți analiza acestui model prin tehnica servirii de tip fază pe care am prezentat-o în capitolul anterior, cel puțin pentru cazul stării stabile. În orice moment considerat, starea modelului de așteptare poate fi definită de numărul existent al clienților, n, plus c numere i_r care specifică faza particulară în care se află clientul din stația r. Intensitatea servirii fiecărei stații este μ, iar a fiecărei faze $k\mu$.

Morse a arătat că deoarece cele c stații de servire sînt identice, nu este nevoie decît să specificăm numărul stațiilor ce au în servire un client în faza r. Astfel, probabilitatea unei stări date poate fi scrisă

$$p(n\,;\, m_1, m_2, \ldots, m_k),$$

m_r este numărul stațiilor ce au în servire un client în faza r. Cititorul va remarca că

$$\Sigma m_r = \begin{cases} n \text{ pentru } n < c, \\ c \text{ pentru } n \geqslant c. \end{cases}$$

Ecuaţiile stării stabile sînt :

$$(1-\delta_{0m_k})\,\lambda p\,(n-1\,;\,m_1,\,\ldots,\,m_k-1)+$$

$$+\,(m_1+1)k\mu\,p(n+1\,;\,m_1+1,\,m_2,\,\ldots,\,m_k)+$$

$$+\sum_{i=1}^{k-1}(1-\delta_{0m_r})\,(m_{r+1}+1)k\mu\,p(n\,;\,m_1,\,\ldots,\,m_r-1,\,m_{r+1}+1,\,\ldots,\,m_k)-$$

$$-\,(\lambda+nk\mu)\,p(n\,;\,m_1,\,\ldots,\,m_k)=0\quad(c>n=\Sigma\,m_r),$$

$$(1-\delta_{0m_k})\,[\lambda p(c-1\,;\,m_1,\,\ldots,\,m_k-1)+$$

$$+\,(m_1+1)\,k\mu p(c+1\,;\,m_1+1,\,m_2,\,\ldots,\,m_k-1)]+ \tag{4.8}$$

$$+\sum_{r=1}^{k-1}(1-\delta_{0m_r})\,(m_{r+1}-1)\,k\mu p\,(c\,;\,m_1,\ldots,\,m_r-1,\,m_{r+1}+1,\,\ldots,\,m_k)-$$

$$-(\lambda+ck\mu)p(n\,;\,m_1,\,\ldots,\,m_k)=0\quad(c=n=\Sigma\,m_r),$$

$$\lambda p\,(n-1\,;\,m_1,\,\ldots,\,m_k)+$$

$$+\,(1-\delta_{0m_k})\,(m_1+1)\,k\mu d(n+1\,;\;m_1+1,\;m_2,\,\ldots,\,m_{k-1},\,m_k-1)+$$

$$+\sum_{r=1}^{k-1}(1-\delta_{0m_r})\,(m_{r+1}+1)\,k\mu p\,(n\,;\,m_1,\,\ldots,\,m_r-1,\,m_{r+1}+1,\,\ldots,\,m_k)-$$

$$-\,(\lambda+ck\mu)p\,(n\,;\,m_1,\,\ldots,\,m_k)=0\;\;(n>c=\Sigma\,m_r).$$

Morse a rezolvat algebric aceste ecuaţii pentru $c=2$, $k=2$. Analiza este excesiv de laborioasă şi se pare că este practic imposibil de a trata astfel cazuri mai ample. Pentru rezolvarea unui sistem pentru care modelul $M/E_k/c$: (∞/FIFO) este cel mai adecvat, este necesar a se aplica unul din următoarele trei procedee :

1. Utilizarea calculatorului electronic pentru a rezolva ecuaţiile stării stabile, pentru cazuri numerice specificate.

2. Simularea procesului (absolut necesară dacă se cer informații legate de stările de tranziție — vezi cap. 6).

3. Utilizarea unui model aproximativ (vezi cap. 8 și 10).

Nu există tabele sau grafice capabile să furnizeze informații legate de timpul total de așteptare sau de timpul de așteptare pentru astfel de modele, exceptînd cazuri foarte rare, pentru care $c = 2$ sau $c = 3$.

4.1.3. $M/D/c : (\infty/\mathrm{FIFO})$

Acest model a fost studiat pentru prima dată în 1930 de către Crommelin. Au fost obținute rezultate doar pentru starea stabilă. Fie p_n probabilitatea existenței a n clienți în sistem, așteptînd sau fiind serviți, și fie a_n probabilitatea cumulată :

$$a_n = \sum_{i=0}^{n} p_i .$$

Putem calcula timpul în multipli ai timpului de servire constant, care el însuși poate fi considerat ca fiind unitatea de timp. Astfel putem lega numărul clienților prezenți la începutul unui interval de timp de cel al clienților de la sfîrșitul intervalului :

$$p_0 = a_c \, \mathrm{e}^{-\lambda}.$$

(Faptul că la începutul intervalului de timp sînt mai puțini clienți decît stații de servire și faptul că în intervalul de timp considerat nu se înregistrează sosiri implică faptul că la sfîrșitul intervalului nu vor fi clienți pentru servit.)

$$p_1 = a_c \, \lambda \mathrm{e}^{-\lambda} + p_{c+1} \, \mathrm{e}^{-\lambda},$$

. . . . . .

$$p_n = a_c \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda} + p_{c+1} \frac{\lambda^{n-1}}{(n-1)!} \mathrm{e}^{-\lambda} + \ldots + p_{n+c} \, \mathrm{e}^{-\lambda}.$$

Cititorul își va reaminti folosirea acestor ecuații, pentru cazul $c = 1$, în capitolul precedent. Acestea sînt ecuațiile de balanță ale lui Crommelin.

Se poate arăta că funcţia generatoare a probabilităţilor p_n, notată $P(s)$, este dată de formula

$$P(s) = -\frac{c-\lambda}{(1-s_1)(1-s_2)\ldots(1-s_{c-1})} \cdot \frac{(s-1)(s-s_1)\ldots(s-s_{c-1}),}{1-s^c\, e^{\lambda(1-s)}}, \tag{4.9}$$

în care $1, s_1, s_2, \ldots, s_{c-1}$ sînt cele c zerouri ale expresiei

$$1 - s^c\, e^{\lambda(1-s)},$$

situate pe cercul unitate din planul complex. Probabilitatea p_n poate fi obţinută prin derivarea succesivă a funcţiei $P(s)$ şi punînd apoi $s = 0$:

$$p_n = \left[\frac{d^n}{ds^n} P(s)\right]_{s=0}.$$

Timpul mediu de aşteptare poate fi dedus calculînd pe $P'(1)$ în maniera obişnuită. Rezultatul obţinut va fi

$$W_q = \frac{1}{\lambda}\sum_{r=1}^{c-1}\frac{1}{1-s_r} + \frac{\lambda^2 - c(1-c)}{2\lambda(c-\lambda)} = \tag{4.10}$$

$$= \sum_{i=1}^{\infty} e^{-i\lambda}\left[\sum_{j=ic}^{\infty}\frac{(i\lambda)^j}{j!} - \frac{c}{\lambda}\sum_{j=ic+1}^{\infty}\frac{(i\lambda)^j}{j!}\right].$$

Calculul nu se poate realiza prea rapid şi de aceea este preferabilă utilizarea unui grafic ce din fericire există pentru W_q cu $1 \leqslant c \leqslant 100$ şi $0{,}10 \leqslant \rho \leqslant 0{,}96$, unde ρ este ca de obicei $\lambda/(c\mu)$.

În multe aplicaţii se cere să aflăm probabilitatea $P(>t)$ ca timpul de aşteptare să depăşească durata t. Aceasta are o expresie foarte complexă şi practic inutilizabilă, exceptînd cazurile în care se utilizează calculatorul electronic. Din fericire, pot fi găsite grafice foarte utile pentru $c = 1, 2, 3, 4, 5, 8, 10, 20$ şi pentru $\rho \leqslant 0{,}8$, care satisfac în mare parte cerinţele practice.

Modelul pe care-l analizăm, $M/D/c$: (∞/FIFO) pare a avea importante aplicaţii în telecomunicaţii (Crommelin a fost unul din conducă-

torii Poștei Centrale Britanice), dar nu am avut niciodată ocazia să-l utilizez în lucrările de cercetare operațională*. De aceea, nu am putut da nici un exemplu practic al acestui model.

4.2. FENOMENE DE DEPLASARE

Modelele cu mai multe șiruri de așteptare discutate pînă acum presupuneau că stațiile de servire sînt alimentate cu clienți de la un singur șir de așteptare. În capitolul anterior am prezentat foarte pe scurt cazul unei singure stații de servire, alimentate de la mai multe șiruri de așteptare, fiecare șir corespunzînd unei anumite priorități de servire. În mod analog, am putea analiza mai multe șiruri de așteptare ce alimentează mai multe stații de servire, conform cu reguli cu priorități. Modelele teoretice corespunzătoare acestui tip de sisteme, create pînă în prezent, sînt puține la număr și neadecvate, iar problemele legate de astfel de sisteme sînt rezolvate cel mai bine prin tehnicile de simulare (vezi cap. 6). Pînă de curînd s-a realizat foarte puțin din punct de vedere teoretic în legătură cu sistemele cu mai multe stații de servire, fiecare avînd un șir de așteptare. Situația se schimbă însă pe zi ce trece. Vom considera în cele ce urmează cîteva noutăți.

Vom presupune că sosirile sînt repartizate Poisson și că, la sosire, fiecare client poate alege așezarea sa într-un șir de așteptare înaintea uneia dintre cele c stații de servire. Este evident că dacă s-ar putea determina regula care să ne spună la ce șir se va așeza clientul și dacă, odată așezat în șir, clientul rămîne acolo, atunci așa-numitul *sistem cu mai multe stații de servire* nu este altceva decît mai multe sisteme cu cîte o singură stație de servire, fără vreo legătură între ele. De exemplu, dacă clienții sînt numărați începînd de la un moment dat și dacă a n-a sosire este alocată unui fir în concordanță cu regula „clienții se vor așeza în șirul numărul n (modulo c)", atunci sistemul va fi format din c procese distincte $E_c/G/1 : (\infty/GD)$, în care nu am definit deocamdată nici repartiția timpului de servire și nici disciplina de așteptare, corespunzătoare fiecărui fir din cele c existente. Sisteme cu o stație de servire de acest fel au fost studiate de mai mulți autori și în special de R. R. P. Jackson și D. Nickols. De acestea însă nu ne vom ocupa în prezenta carte. Vom fi mai interesați de situațiile în

* Afirmația nu mai este adevărată. Modelul se folosește în proiectarea sistemelor de computere.

care se permite clientului să se mute în voie de la un şir la altul, cu intenţia de a-şi micşora timpul de aşteptare. În cap. 10 vom considera un caz al unui astfel de sistem. Pînă atunci, vom examina pe scurt unele cercetări teoretice recente legate de acest fenomen, numit de *deplasare.* O excelentă sinteză* a cercetării acestui fenomen a realizat-o E. Koenigsberg; această sinteză stă la baza prezentei secţiuni. Cititorului care doreşte o analiză mai profundă a acestui subiect îi este recomandată lucrarea lui Koenigsberg.

Strategiile posibile ce se oferă unui client care doreşte să fie servit de la un sistem cu mai multe şiruri de aşteptare sînt foarte numeroase, însă principalele sînt următoarele ($c = 2$):

1. *Regula hotelierului* sau *principiul firului de aşteptare controlat.* Acesta este un principiu general pe care l-am amintit deja în cadrul acestei cărţi, începînd cu prima pagină a capitolului 2, legat de controlul paşapoartelor la aeroporturi. Clientul procedează în felul următor: dacă ambele staţii sînt ocupate, el va aştepta în singurul şir şi se va deplasa conform regulii FIFO. Dacă ambele staţii sînt libere, el are posibilitatea de a alege oricare, ambele staţii putînd fi alese cu probabilităţi egale.

2. *Regula Krishnamoorthi.* Dacă ambele staţii sînt libere, clientul va alege prima cu probabilitatea p_1 şi a doua cu probabilitatea p_2, în rest totul este la fel ca la regula hotelierului.

3. *Regula ghişeelor CEC*: fiecare staţie de servire are propriul său şir de aşteptare. La sosire clientul se va aşeza la cel mai scurt şir. Dacă ambele şiruri sînt libere, va alege pe oricare dintre ele, cu probabilităţi egale. Odată intrat în şir, clientul va rămîne în el. Vom nota această regulă a disciplinei de aşteptare prin EK—1, ceea ce înseamnă de fapt prima regulă a lui Koenigsberg. Este natural să codificăm această regulă ca şi cele ce urmează, cu iniţialele persoanei care le-a clasificat pentru prima dată.

4. *Regula ghişeelor CEC cu deplasare instantanee* sau EK—2: este asemănătoare lui EK—1, dar clienţii se pot deplasa de la un şir la unul mai scurt. Deplasarea în acest caz poate fi aleatoare sau instantanee. Deci avem două cazuri EK — 2a şi respectiv EK—2b. Pentru ambele, principiul general al operaţiei este acelaşi ca şi la EK—1, cu o singură modificare. În cazul EK—2 a, clienţii părăsesc şirul mai lung pentru unul mai scurt cu o viteză proporţională cu diferenţa de lungime dintre ele. În EK—2 b, în momentul în care

* Koenigsberg. E. *On Jockeying in Queues.* CORS—ORSA Joint National Conference, Montreal, 1964.

diferența lungimilor depășește unitatea, ultimul client din șirul mai lung se va deplasa imediat la sfîrșitul celui mai scurt.

5. *Schimbarea cărării* sau EK—3; clienții nou sosiți aleg un șir inițial, dar se deplasează de la șirul mai lung spre cel mai scurt cu o viteză proporțională cu diferența de lungime dintre cele două șiruri. Un client particular aflat într-un șir mai lung poate sau nu să se deplaseze, deoarece ceilalți pot face deplasarea înaintea lui și astfel șirurile se vor egaliza înainte ca el să se poată mișca. (Cu alte cuvinte, transferurile de la un șir la altul au loc în mod continuu și nu sînt simultane.)

6. *Schimbarea drumului* sau EK—4; clienții aleg un șir inițial dar se permite deplasarea în ambele sensuri. Probabilitatea părăsirii unui șir în favoarea altuia este dependentă numai de lungimea lui. Clientul nu are nici un fel de informație legată de lungimea șirurilor, exceptînd cea a șirului în care se găsește.

Koenigsberg a examinat toate aceste principii pentru modelul $M/M/2 : (\infty/\mathrm{GD})$, GD fiind de fapt înlocuit de fiecare regulă în parte. În acest fel, el a demonstrat că fenomenul de deplasare contribuie la îmbunătățirea performanței sistemului de așteptare doar dacă se reduce cantitatea de timp liber al stațiilor. Este singurul aspect meritoriu al fenomenului de deplasare.

Nu vom acorda prea multă atenție așa-numitei reguli a hotelierului; cu toate că aceasta este o procedură mai eficientă decît cea asociată regulii ghișeelor CEC sau EK—1.

Formulele pentru $M/M/c : (L/\mathrm{FIFO})$ pe care le-am dedus anterior demonstrează din plin acest lucru. De exemplu, timpii totali de așteptare rezultați dintr-un sistem $M/M/c : (\infty/\mathrm{FIFO})$, în care intensitatea sosirilor este λ, sînt mai scăzuți decît cei ale celor c sisteme independente de forma $M/M/1 : (\infty/\mathrm{FIFO})$ în care pentru fiecare stație $\lambda' = \lambda/c$. Principiul lui Krishnamoorthi ar putea fi utilizat, conform părerii autorului său, atunci cînd intensitățile de servire ale stațiilor sînt diferite, iar clientul care sosește și găsește ambele stații libere se va duce întotdeauna la cea mai rapidă. Aceasta pare a fi o idee interesantă: Krishnamoorthi a demonstrat că aceasta minimizează numărul mediu al clienților din sistem; din păcate însă, nu poate fi aplicat în multe situații reale, deoarece clientul nu are posibilitatea de a cunoaște care este stația cea mai rapidă. În mod ocazional este posibil ca stațiile de servire să îndrume ele însele un client nou sosit la stația pe care ei o cunosc ca fiind cea mai avantajoasă, iar cînd acest lucru are loc, regula conferă avantajele prezentate. O asemenea situație poate fi remarcată în procesele legate de operațiile

industriale, în care stațiile sînt de fapt utilaje cu viteze de lucru diferite. Aceasta a fost situația reală într-un caz descris de noi, dintr-un alt punct de vedere, însă, într-o lucrare anterioară.

Faptul cel mai interesant legat de modelul $M/M/2:(\infty/\mathrm{EK}-1)$, în care intensitățile medii de servire ale celor două stații sînt μ_1 și respectiv μ_2, este că lungimile medii ale celor două șiruri de așteptare sînt aceleași și sînt independente de intensitățile de servire. De aceea, preferințele clientului nu afectează lungimile medii ale șirurilor atunci cînd ele sînt de forma „alege cel mai scurt șir". Dar clienții care se așază la șirul cu servirea cea mai lentă vor trebui să aștepte în medie mai mult decît ceilalți. Proporția clienților în fiecare șir este

$$\frac{\mu_i}{\mu_1 + \mu_2} \qquad (i = 1, 2). \tag{4.11}$$

Deplasarea probabilistică în cazul sistemului în care se aplică regula ghișeelor CEC nu servește unor scopuri utile. Modelul de așteptare prezintă aceleași caracteristici privind lungimea șirului ca și cînd fiecare client s-ar așeza în șirul cel mai scurt și ar rămîne în el, iar acest rezultat nu este deloc afectat de valorile constantelor de proporționalitate ce măsoară intensitatea deplasării. Această concluzie apare și în cazul EK—5, sau al procedurii schimbării cărării. Acestea două sînt proceduri fără utilitate și ambele determină o anumită pierdere a timpului potențial de servire, care are ca urmare o scădere a capacității de servire. Acest lucru poate fi remarcat prin observația că stările $(n_1, 0)$ și $(0, n_2)$ sînt posibile pentru toți n_1 și n_2, în ambele cazuri, unde n_1 și n_2 sînt numerele clienților aflați în primul și respectiv în al doilea șir.

Modelul $M/M/2:(\infty/\mathrm{EK}-2\mathrm{b})$ are aceleași caracteristici ale lungimii medii a șirului ca și cunoscutul model $M/M/2:(\infty/\mathrm{FIFO})$; acest lucru este adevărat chiar și atunci cînd intensitățile medii de servire diferă; altfel spus, rezultatul se menține pentru intensitățile medii de servire μ_1 și μ_2. Rezultatele modelelor $M/M/c:(\infty/\mathrm{FIFO})$ cu intensitatea de servire μ_i pentru stația i au fost obținute de Gumbel.

Koenigsberg a limitat studiul său la cazul a două stații de servire. Atunci cînd există mai multe stații, să spunem, mai mult de cinci, încep să apară alte efecte ale comportamentului clienților. Pentru astfel de cazuri nu prea există modele matematice aplicabile și aceasta datorită faptului că este dificil a determina exact comportamentul clienților. În cap. 10 vom întîlni totuși un proces de acest tip.

5

MODELE ALE SISTEMELOR SERIALE ŞI CICLICE

Multe dintre sistemele de aşteptare existente în societate nu sînt formate doar dintr-o singură fază a procesului de servire. Este foarte frecvent cazul sistemelor în care ieşirile unui proces de aşteptare reprezintă intrările altuia. Acest lucru este ilustrat în fig. 5.1.

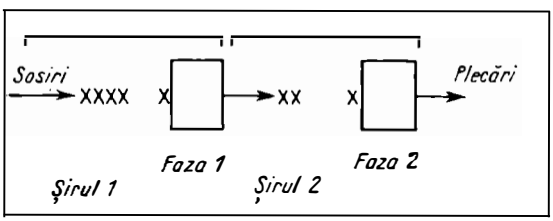


Fig. 5.1. Ilustrarea unui sistem serial în două faze.

Acestea sînt cunoscute sub denumire de sisteme seriale de aşteptare sau sisteme cu staţii de servire în serie. Cînd capetele se înlănţuie şi formează o buclă, astfel încît clienţii se vor roti în sistem, sistemul se va numi sistem de aşteptare ciclic.

În cadrul acestui capitol vom analiza sumar modele ale acestor două tipuri speciale de sisteme de aşteptare. Faptul că expunerea este scurtă nu se datoreşte deloc importanţei sistemelor seriale şi ciclice, ci doar lipsei unor teorii bine fundamentate legate de aceste sisteme şi rarităţii exemplelor de aplicabilitate în literatura cercetării operaţionale.

5.1. SISTEME SERIALE

Singurul model al sistemelor seriale care a fost analizat de către matematicieni este acela care conţine un sistem $M/M/c_1 : (\infty/\text{FIFO})$ ce alimentează un alt sistem $M/M/c_2 : (\infty/\text{FIFO})$. J. R. Jackson a împins cercetarea acestui tip de model la limita sa extremă, rezolvîndu-l pentru cazul general care presupune existenţa a k faze distincte. Mulţi alţi cercetători au studiat cazuri particulare, ca de exemplu pentru $c_1 = 1$ sau $c_2 = 1$ etc. Deoarece analiza lui Jackson nu elucidează doar cazul general, ci determină şi formule realmente utilizabile, ni se pare natural să începem prin a-l prezenta. Presupunem că există k faze de servire, cu şiruri nelimitate admise înaintea fiecărei faze, iar sosirile le considerăm aleatoare, cu intensitatea λ pentru prima fază. Faza r conţine c_r staţii de servire, toate avînd repartiţia timpului de servire exponenţială cu media $1/\mu$. Aceasta implică [v. cap. 4, modelul $M/M/c : (L/\text{FIFO})$] că dacă în faza r sînt n_r clienţi, atunci probabilitatea unei plecări într-o perioadă $\mathrm{d}t$ este

$$n_r \mu_r \, \mathrm{d}t + o(\mathrm{d}t), \qquad n_r < c_r,$$

$$c_r \mu_r \, \mathrm{d}t + o(\mathrm{d}t), \qquad n_r \geqslant c_r.$$

Jackson a dedus soluţia stării stabile a acestui model. Sarcina determinării soluţiilor dependente de timp este considerabilă (şi, din punct de vedere practic, neesenţială). Ecuaţiile stării stabile, conform notaţiilor lui Jackson, sînt următoarele :

$$\left[\lambda + \sum_{r=1}^{k} \delta(n_r) \, a(n_r) \, \mu_r \right] p(n_1, n_2, \ldots, n_k) =$$

$$= \sum_{r=1}^{k} \delta(n_r + 1) \, a(n_r + 1) \, \mu_r \, p(n_1, \ldots, n_r + 1, n_{r+1} - 1, \ldots, n_k) +$$

$$+ \lambda p \, (n_1 - 1, n_2, n_3, \ldots, n_k). \tag{5.1}$$

Trebuie subînţeles faptul că dacă un argument de forma $n_{r+1} - 1$, de exemplu, devine negativ, atunci probabilitatea corespunzătoare poate fi ignorată, fiind nulă. Mai mult, termenul cu -1 şi acela cu $+1$ pot fi neglijaţi în termenul final al ultimei sume.

Simbolurile speciale folosite au următoarele semnificaţii :

$$a(n_r) = \begin{cases} n_r, & n_r < c_r, \\ c_r, & n_r \geqslant c_r, \end{cases} \quad r = 1, 2, \ldots, k,$$

$$\delta(n_r) = \begin{cases} 1, & n_r \neq 0, \\ 0, & n_r = 0, \end{cases}$$

$$\delta\,(n_{k+1}) = 1.$$

Aceste ecuaţii sînt destul de dificil de interpretat şi de aceea, înainte de a trece la prezentarea soluţiei lui Jackson, cititorului îi este recomandat să urmărească aceste ecuaţii pentru $k = 2$ faze. Ele iau forma :

$$\lambda p\,(0, 0) = \mu_2\, p(0, 1)\,;$$

$$(\lambda + \mu_1)\, p(1, 0) = \lambda p(0, 0) + \mu_2\, p(1, 1)\,;$$

$$(\lambda + \mu_2)\, p(0, 1) = \mu_1 p(1, 0) + 2\mu_2\, p(0, 2)\,;$$

$$(\lambda + \mu_1 + \mu_2)\, p(1, 1) = \lambda p(0, 1) + \mu_1 p(2, 0) + 2\mu_2 p(1, 2)\,;$$

$$\begin{aligned} (\lambda + n_1\mu_1 + n_2\mu_2)\, p(n_1, n_2) &= \lambda\, p(n_1 - 1, n_2) + \\ &+ (n_1 + 1)\, \mu_1\, p(n_1 + 1, n_2 - 1) + \\ &+ (n_2 + 1)\, \mu_2 p(n_1, n_2 + 1) \qquad (1 < n_1 < c_1\,;\ 1 < n_2 < c_2)\,; \end{aligned} \tag{5.2}$$

$$\begin{aligned} (\lambda + c_1\mu_1 + n_2\mu_2)\, p(n_1, n_2) &= \lambda p(n_1 - 1, n_2) + c_1\, \mu_1\, p(n_1 + 1, n_2 - 1) + \\ &+ (n_2 + 1)\mu_2 p(n_1, n_2 + 1) \qquad (n_1 \geqslant c_1\,;\ 1 < n_2 < c_2)\,; \end{aligned}$$

$$\begin{aligned} (\lambda + n_1\mu_1 + c_2\mu_2)\, p(n_1, n_2) &= \lambda p(n_1 - 1, n_2) + \\ &+ (n_1 + 1)\, \mu_1 p(n_1 + 1, n_2 - 1) + \\ &+ c_2\mu_2\, p(n_1, n_2 + 1) \qquad (1 < n_1 < c_1\,;\ n_2 \geqslant c_2)\,; \end{aligned}$$

$$\begin{aligned} (\lambda + c_1\mu_1 + c_2\mu_2)\, p(n_1, n_2) &= \lambda p(n_1 - 1, n_2) + c_1\mu_1\, p(n_1 + 1, n_2 - 1) + \\ &+ c_2\mu_2 p(n_1, n_2 + 1) \qquad (n \geqslant c_1\,;\ n_2 \geqslant c_2). \end{aligned}$$

Aceste ecuații sînt într-adevăr destul de dificile. Totuși soluțiile pentru $p(n_1, n_2)$ sînt relativ simple. Să ne întoarcem însă la modelul general al lui Jackson, cu k faze. Soluția este

$$p(n_1, n_2, \ldots, n_k) = p(0, 0, \ldots, 0) \prod_{r=1}^{k} b(n_r), \tag{5.3}$$

în care

$$b(n_r) = \begin{cases} \dfrac{1}{n_r!}(c_r\ \rho_r)^{n_r}, & n_r < c_r\ ; \\ \dfrac{c_r^{c_r}}{c_r!}\ \rho_r^{n_r}\ , & n_r \geqslant c_r\ . \end{cases}$$

Cititorul care va face o comparație cu capitolul anterior va găsi aceste expresii asemănătoare cu soluțiile lui p_n pentru modelul $M/M/c$: (∞/FIFO).

Termenul $p(0, 0, \ldots, 0)$ se determină prin însumare, după n, deoarece suma tuturor termenilor este egală cu 1. Deci

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_k=0}^{\infty} \left[\prod_{r=1}^{k} b\,(n_r) \right] =$$

$$= \prod_{r=1}^{k} \left[\sum_{j=0}^{\infty} b(n_j) \right] = \prod_{j=1}^{k} A_j \quad (j = 1, 2, \ldots, k). \tag{5.4}$$

Din nou, o vizualizare a formei acestei soluții pentru $k = 2$ ar putea fi utilă. De exemplu, pentru $n_1 < c_1$ și $n_2 < c_2$, putem scrie

$$p(n_1, n_2) = \tag{5.5}$$

$$= \frac{\dfrac{1}{n_1!}(c_1\ \rho_1)^{n_1}}{\displaystyle\sum_{r=1}^{c_1-1} \frac{(c_1\ \rho_1)^r}{r!} + \frac{c_1^{c_1}}{c_1!} \cdot \frac{1}{1-\rho_1}} \cdot \frac{\dfrac{1}{n_2!}(c_2\ \rho_2)^{n_1}}{\displaystyle\sum_{s=0}^{c_2-1} \frac{(c_2\ \rho_2)^s}{s!} + \frac{c_2^{c_2}}{c_2!} \cdot \frac{1}{1-\rho_2}} \cdot$$

Formele pentru celelalte combinaţii ale lui n în funcţie de c sînt analoage. Evident că numitorul este egal cu $[p\ (0, 0)]^{-1}$.

Soluţia obţinută poate fi scrisă de asemenea şi sub forma

$$p\ (n_1, n_2) = p^{(1)}\ (n_1) \cdot p^{(2)}\ (n_2), \tag{5.6}$$

în care $p^{(j)}(n_j)$ pentru $j = 1$ sau 2 sînt soluţiile corespunzătoare pentru cele două modele de aşteptare $M(\lambda)/M(\mu_1)/c_1 : (\infty/\mathrm{GD})$ şi respectiv $M(\lambda)/M(\mu_2)/c_2 : (\infty/\mathrm{GD})$, după cum se constată imediat (vezi capitolul anterior), adică soluţia găsită cu mulţi ani în urmă de către Karl Erlang pentru modelele cu o singură fază.

Ca implicaţie a acestei situaţii, ieşirile primei faze trebuie să fie Poisson cu media λ. Astfel spus, trebuie să fie, din punct de vedere statistic, de aceeaşi formă ca şi intrările. Acest lucru va permite cu siguranţă celor două faze să se comporte ca două şiruri independente. De fapt, ieşirile modelului $M/M/c : (\infty/\mathrm{GD})$ sînt aleatoare, mai precis Poisson, cu aceeaşi medie ca şi intrările. Există o demonstraţie foarte riguroasă a acestei informaţii într-un articol al lui Burke, pe care însă nu o vom prezenta aici. Este absolut necesar a sublinia utilitatea practică a acestui fapt : dacă nu ar fi fost adevărat, atunci multe sisteme de aşteptare care pot fi studiate (cel puţin la un nivel rezonabil de aproximare) cu ajutorul modelelor independente ale sistemelor mai simple ar fi, din punct de vedere matematic, de nerezolvat. Simularea pe calculatorul electronic ar deveni în acest caz absolut necesară. Cititorul poate observa că se pot petrece lucruri foarte stranii în cazul în care capacitatea şirului primei faze a unui model serial, să zicem cu două faze, este limitată; de exemplu, dacă prima fază este reprezentată de modelul $M/M/c_1 : (L/\mathrm{GD})$. De fapt, o anumită proporţie a traficului primei faze, cu o intensitate de $\lambda(c_1\mu_1)$ este înlăturată, datorită sosirii unor clienţi atunci cînd capacitatea sistemului, L, este completată. De aceea, traficul care trece prin a doua fază este oarecum diminuat. Problema care se pune este următoarea : care este repartiţia intervalelor dintre plecările succesive care constituie ieşirile filtrate de la prima fază?

Ca şi în multe alte cazuri, răspunsul la această întrebare poate fi dat prin intermediul următorului exemplu forarte simplu. Să considerăm modelul $M/M/1 : (1/\mathrm{GD})$ — disciplina de aşteptare este de fapt nesemnificativă pentru această situaţie — în care sosirile au loc cu o intensitate medie λ iar timpul mediu de servire este $1/\mu_1$. Este limpede că atunci cînd un client este în curs de servire, toate celelalte sosiri sînt respinse. Prin urmare, intervalele dintre plecările

consecutive din sistem, aşa-numiţii *timpi dintre plecări*, sînt fiecare suma a cîte două componente :

t_1 — timpul scurs între o plecare din sistem şi o nouă sosire în sistem ;

t_2 — timpul necesar servirii noului sosit.

Deoarece repartiţia timpilor dintre sosiri este exponenţială, t_1, prima din aceste componente, este total independentă de momentul în care clientul anterior — care poate a fost respins — a sosit, ca şi de momentul celei mai recente plecări. (Reamintiţi-vă „uitarea" caracteristică repartiţiei exponenţiale, menţionată în cap. 2.) Prin urmare, din punct de vedere statistic, prima componentă are aceeaşi repatriţie cu cea a timpilor dintre sosiri, adică este exponenţială cu media $1/\lambda$. A doua, prin definiţie, este de asemenea repartizată exponenţial.

Ca urmare, timpii dintre plecările din sistem sînt repartizaţi ca suma a două variabile exponenţiale, astfel încît repartiţia lor nu poate fi exponenţială. Acest raţionament referitor la faptul că ieşirile dintr-un proces $M/M/1$: (1/GD) nu sînt Poisson a fost făcută de către P. G. Finch, care a demonstrat că :

1. Intervalele dintre plecări sînt repartizate exponenţial *doar atunci* cînd capacitatea sistemului este infinită sau, utilizînd notaţiile noastre , $L = \infty$.

2. Mai general, o capacitate infinită a sistemului ($L = \infty$) şi o repartiţie exponenţială a timpilor de servire sînt condiţii necesare şi suficiente pentru ca intervalele dintre plecări să fie independente.

De aceea, dacă aşteptarea din prima fază a unui proces format din două faze este limitată, atunci cele două faze nu sînt independente, chiar dacă între faze este permisă o aşteptare nelimitată (acest rezultat se menţine şi în cazul existenţei mai multor şiruri paralele, pentru fiecare fază). Intrările în cea de a doua fază nu sînt repartizate Poisson.

În discuţia ce urmează, despre fenomenul de blocare, vom arăta că cele două faze nu sînt independente dacă reciproca este adevărată ; adică, în caz că avem o capacitate infinită a sistemului în prima fază şi o capacitate limitată pentru aşteptarea dintre faze. Este important să subliniem că restricţiile făcute capacităţii sistemului sau, altfel spus, a spaţiului de aşteptare la orice fază a unui sistem serial multiplu, determină dependenţa fazelor. În practică însă, de obicei se acceptă ipoteza că fazele unui sistem, să spunem, bifazic, sînt independente, în cazul în care este admisă o capacitate suficient de mare înaintea fiecărei faze ; o metodă practică de a decide ce înseamnă o capacitate suficient de mare este aceea de a calcula probabilitatea de

refuzare a unei sosiri. O probabilitate mai mică decît 0,10 înseamnă practic capacitate suficientă. Cu cît este mai mică probabilitatea refuzului unei sosiri, cu atît încrederea în independenţa fazelor este mai mare. (În cap. 15 vom face referiri mai largi legate de aceste aspecte.)

Să ne întoarcem la sistemul de aşteptare $M/M/1$: (1/GD). Ar fi util să ni-l închipuim ca fiind prima fază a unui sistem bifazic, în care faza a doua este formată dintr-o staţie exponenţială cu timpul mediu de servire $1/\mu_2$, cu o capacitate infinită, şi să ne imaginăm condiţiile pentru care şi acest sistem simplu poate fi considerat ca fiind format din două faze independente. Există două cazuri speciale de acest fel, ambele extreme. Primul corespunde unei intensităţi a traficului $\rho = \lambda/\mu_1$ foarte scăzute pentru prima fază şi uneia moderat de mare pentru a doua. Deci nu are loc aproape nici o respingere a vreunui client în prima fază şi, cum nici unul nu va fi nevoit să aştepte, repartiţia timpului dintre plecări din prima fază este identică, din punct de vedere statistic, cu repartiţia timpului dintre sosiri. Efectul primei faze de servire este acela de a determina timpi exponenţiali de sosire şi nu de a modifica repartiţia timpilor dintre sosiri. În acest caz, faza secundă poate fi tratată ca un sistem $M/M/1$: (∞/GD). Astfel de situaţii, de obicei, au mai multe staţii paralele pentru fiecare fază, pot fi uneori întîlnite cînd prima fază este o operaţie de servire de genul „agentului de circulaţie". Exemple de acest gen apar în telecomunicaţii şi în restaurante.

Al doilea caz special corespunde suprasaturării primei faze, adică cazului în care intensitatea traficului depăşeşte unitatea. În ciuda proporţiei ridicate a respingerilor, prima fază este aproape permanent ocupată, astfel încît repartiţia intervalelor dintre plecări este, din punct de vedere statistic, aproape identică cu repartiţia timpului de servire. Dacă aceasta din urmă este, aşa cum presupunem, exponenţială, înseamnă că faza a doua are intrările Poisson, şi poate fi considerată independentă de prima. Asemenea situaţii pot fi remarcate în unele birouri de la companiile aeriene, în care apelurile telefonice au o frecvenţă maximă pentru perioadele în care vremea este foarte nefavorabilă.

Un alt aspect interesant al proceselor seriale este fenomenul de *blocare*. Acesta are loc doar atunci cînd se admite un şir de aşteptare limitat în faza a doua, iar clienţii nu pot părăsi prima fază fără a trece prin a doua.

De obicei, lucrările despre teoria așteptării acceptă restricția prin care nu este admis nici un șir de așteptare în faza a doua, dar această simplificare, cu toate avantajele ce le prezintă pentru calcule, nu este necesară. De exemplu, să presupunem, că fiecare fază este formată dintr-un șir individual de așteptare și a doua fază poate cuprinde doar L_2 clienți; în momentul cînd faza a doua este completată, iar un client a fost servit în prima fază, prima fază este blocată. În aceste condiții, valorile date lui L_2 apar fără însemnătate. Dar cu cît L_2 este mai mare, cu atît probabilitatea ca blocarea să se producă este mai mică. Uneori se presupune capacitatea primei faze L_1 ca fiind egală cu 1, astfel încît dacă un client sosește și o găsește ocupată, acesta să fie respins. Această presupunere poate fi însă extinsă și pentru alte valori ale lui L_1. Totuși, cu cît L_1 este mai mare, cu atît respingerile vor fi mai puține.

Pentru exemplificare, vom analiza cazul convențional a două faze exponențiale, de capacitate infinită în prima fază și 1 în a doua. Clientul care dorește să părăsească prima fază pentru a intra într-a doua, în momentul cînd stația corespunzătoare fazei secunde este completă, va trebui să rămînă blocat în prima fază.

Vom defini următoarele probabilități ale stării stabile:

$p(n, 0)$ — există n clienți în prima fază, unul care este servit și $n-1$ așteptînd; nimeni nu se găsește în faza a doua;

$p(n, 1)$ — n clienți în prima fază, unul care este servit și $n-1$ așteptînd; faza a doua completă;

$p_b(n, 1)$ — n clienți în prima fază, unul care a fost servit și blochează stația de servire, în timp ce $n-1$ clienți așteaptă.

Vom presupune că $p_b(n, 1)$ se schimbă instantaneu în $p(n-1, 1)$ în momentul eliberării fazei secunde. Ecuațiile stării stabile sînt:

$$\begin{aligned}
\lambda p(0, 0) &= \mu_2 p(0, 1);\\
(\lambda + \mu_2) p(0, 1) &= \mu_1 p(1, 0) + \mu_2 p_b(1, 1);\\
(\lambda + \mu_2) p_b(1, 1) &= \mu_1 p(1, 1);\\
(\lambda + \mu_1) p(n, 0) &= \lambda p(n-1, 0) + \mu_2 p(n, 1); \quad n > 0; \qquad (5.10)\\
(\lambda + \mu_1 + \mu_2) p(n, 1) &= \lambda p(n-1, 0) + \mu_1 p(n+1, 0) +\\
&+ \mu_2 p_b(n+1, 1); \quad n > 0;\\
(\lambda + \mu_2) p_b(n, 1) &= \lambda p_b(n-1, 1) + \mu_1 p(n, 1); \quad n > 1.
\end{aligned}$$

Aceste ecuaţii pot fi rezolvate în maniera cunoscută, calculul algebric fiind însă destul de anevoios. Elementul de interes este acela că blocarea ce apare reduce intensitatea de servire efectivă a primei faze sub valoarea μ_1; de aceea, volumul maxim de trafic ce poate fi păstrat în echilibru statistic este mai mic decît λ/μ_1. A fost demonstrat că pentru exemplul dat, acest volum este

$$\varepsilon_1^{max} = \frac{\mu_2(\mu_1 + \mu_2)}{\mu_1^2 + \mu_1\mu_2 + \mu_2^2}. \tag{5.11}$$

În cazul particular, cînd $\mu_1 = \mu_2$, ε_1^{max} devine egal cu 2/3. Este fără îndoială o valoare mai mică decît $1 - \eta$, unde η este o cantitate pozitivă, valoare ce rezultă dintr-o stare stabilă în care nu apare blocarea.

Aceasta indică faptul că surplusul de capacitate dintre două faze a sistemului de aşteptare va permite o utilizare superioară a primei faze. Se poate arăta că un surplus de capacitate de numai o unitate măreşte pe ε_1^{max} la valoarea 3/4.

Modelele seriale nu pot fi uşor analizate, exceptînd cazul în care repartiţiile timpilor de servire sînt exponenţiale. Motivul este că dacă aceste repartiţii nu sînt exponenţiale, atunci intervalele dintre plecările consecutive din oricare fază nu sînt independente. De obicei însă, ele nu sînt exponenţiale. Cele două exemple pe care le-am examinat şi în care şirul de aşteptare a fost limitat fie în prima fază, fie în a doua (cu rezultate diferite), arată că problema se complică foarte mult în astfel de situaţii. Se pare că tehnicile de simulare constituie singura cale de determinare a proprietăţilor unui model de sistem de aşteptare, serial de orice grad de complexitate ar fi.

Este posibil a se realiza unele succese în multe aplicaţii practice făcînd presupunerea că faza a doua a unui sistem, să zicem bifazic, este independentă de prima. Dacă prima acţionează ca un filtru din cauza unei capacităţi limitate, va trebui corectată intensitatea medie a intrărilor în faza a doua.

5.2. SISTEME CICLICE

Cu toate că existau de mulţi ani analize legate de sistemele ciclice de aşteptare — cum ar fi analiza făcută de Jackson şi Taylor — principala contribuţie teoretică a fost adusă de către E. Koenigsberg.

Vom încheia acest capitol printr-o succintă prezentare a modelului lui Koenigsberg*.

Modelul este format din k staţii individuale, legate între ele astfel încît formează un circuit închis, ieşirile ultimei faze constituind intrările primei. Toate staţiile prezintă o repartiţie exponenţială a timpilor de servire, iar intensitatea medie de servire pentru faza i este μ_i. Există N clienţi în circuit şi acest număr nu poate fi nici mărit şi nici micşorat (prin eventuale plecări). Probabilitatea ca în sistem să existe n_i clienţi în faza i (incluzîndu-l, ca de obicei, şi pe cel care este servit), unde $i = 1, 2, \ldots, k$, este notată în echilibru statistic cu $p(n_1, n_2, \ldots, n_k)$.

Ecuaţiile stării stabile sînt

$$\sum_{i=1}^{k} \mu_i \, \delta(n_i) \cdot p(n_1, n_2, \ldots, n_k) = \mu_k \, p(n_1 - 1, n_2, \ldots, n_k + 1) +$$

$$+ \sum_{i=1}^{k-1} \mu_i \, p(n_1, \ldots, n_i + 1, n_{i+1} - 1, \ldots, n_k),$$

unde

$$\sum_{i=1}^{k} n_i = N; \quad \delta(n_i) = \begin{cases} 1, & n_i \neq 0, \\ 0, & n_i = 0, \end{cases} \tag{5.12}$$

şi unde oricare probabilitate $p(n_1, n_2, \ldots, n_k)$ cu cel puţin un argument negativ va fi considerată egală cu zero. (Comparaţi notaţia lui Jackson cu aceasta.) Koenigsberg a demonstrat că soluţia acestor ecuaţii este

$$p(n_1, n_2, \ldots, n_k) = \frac{\lambda_1^{n_1} \lambda_2^{n_2} \ldots \lambda_k^{n_k}}{\sum_\pi \lambda_1^{r_1} \lambda_2^{r_2} \ldots \lambda_k^{r_k}}, \tag{5.13}$$

unde $\lambda_i = 1/\mu_i$ şi $\sum_\pi$ este suma după toate valorile întregi care satisfac relaţia $\sum_{i=1}^{k} r_i = N$.

Se vede imediat că utilizarea staţiei i este

$$U_i = 1 - p_i^{(0)} = 1 - \frac{\sum_{\pi i} \lambda_1^{r_1} \lambda_2^{r_2} \ldots \lambda_k^{r_k}}{\sum_\pi \lambda_1^{r_1} \lambda_2^{r_2} \ldots \lambda_k^{r_k}},$$

* Koenigsberg, E. *Cyclic Queues*. Operat. Res. Quart., 9, nr. 1, 1958.

unde $\sum_{\pi_i}$ este $\sum_{\pi}$ cu $r_i=0$. În cazul în care toate staţiile au aceleaşi repartiţii ale timpului de servire, probabilitatea ca faza i să fie liberă este

$$p_i^{(0)} = \frac{\text{numărul diferitelor combinaţii } \{r_1,\ldots,r_k\} \text{ce au suma } N \text{ ,cu } r_i=0}{\text{numărul diferitelor combinaţii } \{r_1, \ldots, r_k\} \text{ ce au suma } N}$$

Koenigsberg a arătat că această expresie este egală cu

$$\frac{k-1}{N+k-1}. \tag{5.14}$$

Formulele nu sînt foarte uşor de utilizat dacă numărul fazelor este mare şi dacă timpii medii de servire sînt toţi diferiţi în cele k faze. Din păcate, în realitate există multe situaţii ce se aseamănă proceselor ciclice de aşteptare. În momentul scrierii acestei cărţi, un set de tabele cu probabilităţi ale proceselor ciclice se realizează de P. J. Sandiford tabele care vor simplifica fără îndoială mult aplicarea lor.

6

TEHNICI DE SIMULARE

6.1. METODE MONTE CARLO

Problema ce urmează este studiată în mod curent în statistică. Sînt date două variabile aleatoare x și y avînd densitățile de repartiție $f(x)$, respectiv $g(y)$, pe intervalele $0 \leqslant x \leqslant \infty$ și respectiv $0 \leqslant y \leqslant \infty$. Se cere să obținem densitatea de repartiție $h(u)$, cu $u = x + y$. Obținerea acestei densități de repartiție se face prin intermediul transformatelor Laplace :

$$f^*(z) = \int_0^\infty \mathrm{e}^{-zx} f(x)\, \mathrm{d}x,$$

$$g^*(z) = \int_0^\infty \mathrm{e}^{-zy} g(y)\, \mathrm{d}y,$$

$$h^*(z) = f^*(z)\, g^*(z).$$

Astfel $h^*(z)$ este transformata Laplace a lui $h(u)$ și în mod teoretic poate fi inversată pentru a da rezultatul dorit.

Din păcate însă, este foarte dificil sau uneori chiar imposibil să inversăm pe $h^*(z)$ pentru a obține $h(u)$ în termenii unor funcții cunoscute, pentru care există tabele ; de aceea această procedură nu ne poate ajuta decît rareori. Evident este posibil să deducem momentele repartiției lui u, direct din $h^*(z)$, deoarece coeficientul lui $(1 - z)^r$ $(z^r/r!)$ în seria Taylor-MacLaurin pentru $h^*(z)$ este momentul m_r' în vecinătatea originii (moment centrat). Funcția de repartiție a lui u poate fi aproximată prin egalarea momentelor ei m_r în jurul mediei,

cu cele ale unor repartiţii cunoscute, cum este de exemplu una din familia repartiţiilor Pearson.

Totuşi, această procedură nu poate fi întotdeauna acceptată, iar statisticienii au imaginat metodele de determinare a lui $h(u)$, prin metode de sondaje probabilistice sau aşa-numitele *tehnici Monte Carlo.* Aplicarea unor astfel de metode va fi ilustrată prin intermediul unui exemplu.

Să presupunem că un avion trebuie să decoleze din Toronto la ora 09.00. Experienţa anterioară ne spune că decolările la Toronto întîrzie după cum urmează :

Întîrzieri la decolare în minute	0	1	2	3	4	5
Procentul zborurilor întîrziate	60	15	10	7	5	3

Mai mult decît atît, experienţa unor zboruri similare a arătat că durata zborului de la Toronto la Montreal variază de la zbor la zbor astfel :

Durata zborului în minute	58	59	60	61	62	63	64
Procentul zborurilor avînd această durată	10	20	30	20	10	6	4

Dacă s-ar programa sosirea avionului la ora 10.00, cîte avioane (în procente) vor sosi mai tîrziu?

Utilizînd tehnicile Monte Carlo, putem începe prin a repartiza cele 100 de numere 00, 01, 02, ..., 99, întîrzierilor de la decolare şi duratelor zborurilor în concordanţă cu frecvenţele lor, în felul următor :

Întîrzierea la decolare în minute	0	1	2	3	4	5
Numere repartizate acestei întîrzieri	00—59	60—74	75—84	85—91	92—96	97—99

Durata zborului în minute	58	59	60	61	62	63	64
Numere repartizate	00—09	10—29	30—59	60—79	80—89	90—95	96—99

Să ne uităm într-un tabel de numere aleatoare, de exemplu în cel prezentat în cartea lui Fischer şi Yates. Să luăm primele două, care sînt 03 şi 97. Raportîndu-ne la tabelul întîrzierilor la decolare, numărul 03 corespunde unei întîrzieri de 0 minute; iar în tabelul duratelor zborurilor, 97 corespunde unei durate de 64 de minute. În acest caz, ora de sosire va fi 9,00 + 0,64 = 10,04.

Repetînd această procedură de 100 de ori, vom determina numărul zborurilor întîrziate. Acest număr este 59. Prin urmare, metoda Monte Carlo, bazată pe simularea a 100 de zboruri, estimează la 59 % numărul zborurilor întîrziate. Cele 100 de perechi de numere aleatoare alese determină parţial răspunsul la problema pusă. Dar ce s-ar întîmpla dacă am alege alte 100 perechi de numere aleatoare şi am repeta simularea cu ele? Un al doilea experiment Monte Carlo, bazat pe alte 100 de perechi de numere aleatoare arată că 56 % din zboruri vor întîrzia. Există, aşadar, două estimaţii : 59 % şi 56%. Care din ele este corectă? Nici una, deoarece ambele estimează o aceeaşi problemă, al cărui răspuns nu este cunoscut. O estimaţie Monte Carlo bazată pe simularea a 200 de zboruri este mai corectă decît cea bazată pe 100. Cea mai bună estimare a noastră este că

$$\frac{1}{2}(59 + 56) = 57{,}5\%$$

din zboruri vor întîrzia.

De fapt, această problemă poate fi rezolvată riguros prin metode statistice obişnuite. Răspunsul corect este că 58,5 % din zboruri vor întîrzia. Remarcăm că utilizînd metoda Monte Carlo nu vom obţine valoarea reală şi că, utilizînd experimente repetate, va trebui să determinăm şi dispersia valorii determinate. Dacă însă am utiliza metoda transformatei Laplace, în aparenţă aceste operaţii nu ar mai fi necesare. Cu toate acestea ele sînt necesare, deoarece în situaţiile reale sîntem puşi în faţa unor probleme de determinare a dispersiei estimatorilor noştri pentru parametrii lui $f(x)$ şi $g(y)$, în mod necesar tot prin operaţii de sondaj în lumea reală.

Despre tehnicile ce folosesc selecţii aleatoare se pot spune mult mai multe, iar cititorului care doreşte să intre în detaliile acestor probleme îi recomandăm a studia una din cărţile de simulare. Pentru scopul acestei cărţi este suficient a observa că tehnicile de simulare oferă posibilitatea studierii modelelor sistemelor de aşteptare ce nu pot fi tratate analitic. În general o soluţie analitică, datorită generali-

tăţii ei, este preferabilă unei simulări ce este limitată doar la un caz particular şi este valabilă doar la un moment dat, dar fără îndoială că o soluţie determinată pe baza tehnicilor de simulare este mai bună decît nimic.

Realizarea simulării cu mijloace manuale este de foarte lungă durată. Fiecare iteraţie, sau experiment, este în sine o operaţie laborioasă, iar pentru a estima dispersia sînt necesare mai multe iteraţii. Ca urmare, este normal ca simularea să fie programată pe calculatorul electronic. Pentru a realiza acest lucru este necesar :

1. Un mijloc adecvat de generare a numerelor aleatoare.
2. O anumită modalitate de a produce valori avînd repartiţia specificată.
3. Tehnici eficiente de programare.

În cadrul acestui capitol nu este posibil să discutăm aceste probleme în profunzime, dar fiecare dintre ele va fi analizată succint.

6.2. NUMERE ALEATOARE ŞI PSEUDOALEATOARE

Numerele aleatoare utilizate în procesele de simulare trebuie să fie generate la momentul oportun, deoarece în memoria internă a calculatoarelor nu poate fi memorat un număr suficient de mare de astfel de numere antecalculate. Acesta este motivul pentru care tabelele de numere aleatoare nu prea mai sînt utilizate. Este posibil de a produce numere aproape aleatoare (pseudoaleatoare) prin intermediul dispozitivelor electronice, dar acest procedeu nu este un procedeu uzual. În cercetarea practică a proceselor de aşteptare, prin intermediul modelelor, de obicei ne interesează să comparăm efectele schimbării numărului staţiilor, sau a diferitelor reguli de prioritate. Dacă în simulările făcute am utiliza numere aleatoare veritabile, aceste comparaţii vor da rezultate confuze pentru diferitele secvenţe de numere aleatoare utilizate. Ar fi mult de cîştigat, dacă am putea genera o secvenţă particulară după voie. O astfel de secvenţă nu poate fi, prin definiţie, formată din numere aleatoare, dar poate fi aproximativ (suficient de) aleatoare. Asemenea secvenţe se numesc *şiruri de numere pseudoaleatoare*. S-a demonstrat că numerele pseudoaleatoare generate de reguli deterministe (cum dealtfel sînt toate) trebuie să fie ciclice. Un obiectiv important al simulării este de a

asigura cicluri adecvate suficient de lungi. Există mai multe metode de generare a secvențelor pseudoaleatoare. O procedură foarte frecvent utilizată, ce poate genera secvențe cu cicluri lungi, este procedeul multiplicativ congruențial, sugerat pentru prima dată de către Lehmer în 1949 și analizat în detaliu de Moshman în 1954.

Secvența de numere întregi generată de formule de recurență $x_{n+1} = px_n (\text{mod } q)$, în care x_0 este un număr aleator inițial, constituie începutul procesului de simulare. Factorul p este ales astfel încît maximizează lungimea ciclului. În multe situații, numărul q este ales ca fiind cel mai mare întreg pozitiv reprezentabil în calculator, dar dacă calculatorul admite înregistrări de lungime variabilă, lui q i se poate da orice valoare convenabilă.

Problema alegerii valorilor lui p și q care vor produce cicluri lungi este complicată, deoarece ea depinde de caracteristicile calculatorului cu care se lucrează. Operația de congruență implică împărțirea. Aceasta este de obicei cea mai lungă dintre operațiile aritmetice de bază; în scopul de a o scurta cît mai mult posibil, q se ia de obicei egal cu numărul întreg maxim ce se poate reprezenta în calculator. Două combinații, ce au fost foarte mult utilizate, sînt:

$$p = 3^{17}, \qquad q = 2^9,$$

$$p = 13^{13}, \qquad q = 2^{31} - 1.$$

Se întîmplă uneori ca pe parcursul unui proces de simulare să fie utilizate atît de multe numere pseudoaleatoare, încît să depășească numărul celor ce se pot obține cu o formulă congruențială. Este evident că acest lucru este nedorit, deoarece pot să apară corelații „interesante". Această dificultate poate fi depășită prin generarea mai multor șiruri de numere pseudoaleatoare cu ajutorul unor numere x_0 diferite între ele. Utilizînd unul dintre șiruri, determinăm la fiecare fază a procesului de simulare secvența din care va fi ales numărul pseudoaleator.

6.3. GENERAREA VARIAȚIILOR ALEATOARE

Următoarea cerință este de a determina mijloacele necesare pentru a transforma numărul pseudoaleator generat într-o valoare a unei repartiții. Dacă tipul repartiției se identifică foarte exact, în

termeni matematici, atunci există mai multe variante de lucru. Ele sînt descrise în lucrările despre simulare; de exemplu, pentru a genera o valoare dintr-o repartiţie exponenţială se poate proceda conform indicaţiilor date de von Neumann. Mai întîi se generează numere pseudoaleatoare, pînă cînd se obţine o secvenţă de tipul

$$x_1 > x_2 > x_3 > \ldots > x_n \leqslant x_{n+1}. \tag{6.1}$$

Atunci cînd acest lucru s-a realizat iar n este impar, primul număr x_1 este acceptat ca o valoare repartizată exponenţial, cu media egală cu unitatea. Astfel, valoarea căutată se determină înmulţind pe x_1 cu valoarea medie a repartiţiei. Dacă n este par, secvenţa este înlăturată şi începe generarea unei noi secvenţe; dacă aceasta este secvenţa de tipul arătat mai înainte, atunci valoarea căutată este primul număr din această secvenţă x_1', plus 1. Dacă o a treia secvenţă este necesară, al treilea prim număr x_2'' este adunat cu 2 ş. a. m. d. Această procedură generează valori repartizate exponenţial.

Prin ajustarea datelor observate cu o funcţie de repartiţie cunoscută poate să apară un element de aproximaţie care este nejustificat. S-ar putea să avem motive apriori pentru a ajusta datele cu o anumită repartiţie : în acest caz totul este în regulă. Dar nu este întotdeauna aşa. Alternativa este de a reduce datele observate ale repartiţiei la un tabel reprezentînd valorile cumulate ale funcţiei de repartiţie

$$y = F(x).$$

Apoi sînt generate numere aleatoare y, iar tabelul cu $y = F(x)$ va da valorile corespunzătoare x, prin intermediul interpolării. Aceasta este cea mai utilizată metodă în exemplele de simulare ce vor fi prezentate în capitolele următoare.

În sfîrşit, se pune problema tehnicilor eficiente de programare. Cititorul este fără îndoială convins că sarcina scrierii programelor pentru calculator a fost mult uşurată în ultimii ani, datorită dezvoltării limbajelor simbolice de programare, dintre care unele, de exemplu limbajul FORTRAN, este specializat pentru problemele ştiinţifice matematice (în care numărul de calcule este mare, iar volumul intrărilor şi ieşirilor, respectiv al datelor de intrare şi de ieşire este mic) spre deosebire de altele, ca de exemplu limbajul COBOL este specializat pentru prelucrarea datelor economice, admi-

nistrative (unde calculele sînt elementare însă numărul datelor de intrare/ieșire este enorm).

Avantajul acestor limbaje de programare este acela că instrucțiunile transmise calculatorului sînt foarte directe, iar calculatorul translatează aceste instrucțiuni simbolice în limbaj mașină (lucru realizat de către compilator), le alocă memoria necesară și, la nevoie, dispozitive de intrare/ieșire. Odată realizate aceste limbaje de programare, printre analiștii cercetării operaționale a apărut ideea creării unui limbaj simbolic pentru programarea simulărilor. Există mai multe „limbaje de simulare" și printre ele, GSP, Simscript, Simula, EPPSS, și GSL.

Fără îndoială că aceste limbaje ar ușura mult scrierea programelor pentru simulare, odată ce au fost *depășite* dificultățile de stăpînire a limbajului sistemelor de așteptare. Cu toate acestea, există o problemă care se pune și anume aceea că compilatoarele (sau programele de asamblare) pentru aceste limbaje nu pot fi utilizate pentru orice tip de calculator. Se pare că doar limbajul CSL, creat de către Buxton și Laski, a fost implementat pe mai mult de două calculatoare, și numai pe IBM și Honeywell. În plus, pînă acum analistul care are acces la diferite tipuri de IBM (7090, 7094,7044) este singurul care are posibilitatea de alegere a unui limbaj de simulare dintre GPSS, SIMPAC, SIMSCRIPT și CSL. Această limitare poate fi doar momentană, în condițiile în care se pregătește introducerea unei noi generații de calculatoare. În afară de cazul cînd cineva dorește să scrie mai multe programe de simulare în fiecare an, nu cred că merită ca cineva să învețe aceste limbaje sau sisteme care au fost dezvoltate.

O mare parte a simulărilor — și printre ele se numără și cele legate de sistemele de așteptare — ocupă o poziție intermediară între calculul matematic și prelucrarea problemelor economico-administrative. Nu pretind că acest lucru este adevărat și pentru simulările științifice, sau metoda Monte Carlo se aplică, problemelor pur matematice, așa cum apar ele în fizică. Ele pot fi programate pentru scopuri practice în FORTRAN sau COBOL, fără mare pierdere de timp; iar în cazul în care va avea loc schimbarea calculatorului, reprogramarea va fi foarte facil de realizat. Dacă cineva este un mare expert în simulare și dacă există o cît de mică șansă ca întreprinderea la care lucrează să înlocuiască calculatorul existent, ABC 4567 printr-un XYZ 9012, atunci nu numai că este dificil să învețe FAKEITAL sau oricare alt limbaj de simulare considerat ca cel mai eficient la momentul respectiv, dar chiar să creeze un compilator pentru acest nou

calculator special pentru acest limbaj (pe care probabil nimeni altcineva nu-l are). Ar fi foarte interesant de ştiut ordinul de mărime a economiilor care se obţin prin utilizarea acestor limbaje speciale, nu ca o proporţie a timpului sau a costului programării simulării, ci ca o proporţie a costului întregului proiect de cercetare operaţională. (Am încercat să obţin răspunsuri la această chestiune, în timpul unei sesiuni a unei conferinţe pe care am organizat-o în America de Nord. Cei cărora le-am scris dinainte despre acest lucru mi-au răspuns evaziv. Cititorul poate interpreta aceste rînduri cum doreşte.)

6.4. UTILIZAREA SIMULĂRILOR

Odată analizate cerinţele majore ale simulării, trebuie să discutăm două aspecte particulare ale utilizării lor: stabilitatea estimărilor obţinute pe baza lor şi modul optim de planificare a experimentelor de simulare.

Am văzut anterior că un model de aşteptare, de exemplu de tipul $M/M/1 :(\infty/\text{FIFO})$ în care operaţia propriu-zisă este începută dintr-o stare iniţială, la momentul 0, trece printr-o fază intermediară, tranzitorie, înainte de a ajunge în starea stabilă. Acest fenomen se întîlneşte şi în cadrul simulării. Atunci cînd se începe să se simuleze un sistem de aşteptare (sau, pentru a fi mai exacţi, să se studieze un model de aşteptare prin intermediul simulării), vor trebui date anumite valori iniţiale pentru parametrii sistemului; de exemplu, va trebui să ştim numărul clienţilor existenţi în fiecare şir (în caz că sînt mai multe), duratele serviciilor încă neterminate pentru clienţii aflaţi în curs de servire şi altele.

Este evident că aceste condiţii iniţiale vor avea un efect asupra rezultatelor finale determinate pe baza simulării făcute pentru o perioadă de timp. Problema este: care trebuie să fie durata intervalului de timp pentru care se face simularea, astfel încît să se ajungă la o stare de echilibru statistic?

Nu există un răspuns direct, imediat, pentru această problemă. În cap. 15 vom analiza o problemă ce a fost iniţial studiată prin simulare. Această problemă presupune un sistem de aşteptare foarte riguros definit, fără nici un fel de consideraţii ambigue. Acele simulări, în care erau în jur de 20 de staţii de servire, nu au produs estimări stabile acceptabile ale 80-cuantilelor repartiţiei timpului de

aşteptare, pînă cînd nu au fost simulaţi timpii de servire pentru aproximativ 1000 de clienţi. În decursul anilor, au fost propuse fel de fel de scheme pentru a surmonta această dificultate.

Prima şi cea mai elaborată schemă este aceea în care calculatorul realizează estimări la intervale date pentru un sistem de parametri selectaţi ; de exemplu, determinarea probabilităţii ca timpul de aşteptare să fie nul, bazată pe proporţia clienţilor ce nu aşteaptă deloc din grupe succesive de 50 de clienţi, corelată cu proiectarea unor teste care să pună în evidenţă eventualele tendinţe de autocorelare. Putem utiliza în acest sens testul Smirnov-Kolmogorov. Atunci cînd se stabileşte că parametrii nu mai sînt autocorelaţi, rezultatele anterioare sînt eliminate şi se continuă să se simuleze pentru o perioadă următoare, prestabilită. Acest ultim stadiu este considerat a fi „adevărata" simulare. Din păcate, nu ştiu dacă aceasta este o metodă utilizabilă practic sau nu.

A doua metodă, realizată tot cu calculatorul, este de a tipări la anumite intervale estimaţiile unor parametri selectaţi, astfel încît, la finalul fiecărei rulări, valorile succesive să poată fi reprezentate grafic şi prin aceasta să poată fi înlăturate rezultatele obţinute în faza tranzitorie. (Aceasta a fost procedura adoptată în majoritatea simulărilor prezentate în această lucrare, exceptînd cele din cap. 12.) S-a sugerat de asemenea, ca fiecare rulare să simuleze un anumit număr, să spunem 4 sau 5, de perioade ocupate, înainte să ne considerăm în echilibru statistic. Singurul dezavantaj al acestei idei ingenioase este acela că pentru intensităţi mari ale traficului perioadele ocupate sînt suficient de lungi, iar ciclurile de simulare vor deveni foarte costisitoare. De fapt, nu există o „regulă de aur" ce poate fi aplicată ca soluţie a acestei probleme, ci doar o indicaţie : atenţie!

6.5. EXPERIMENTE DE SIMULARE

Ultima problemă majoră ce se pune acum este de a determina varianta optimă de planificare a experimentelor de simulare. Aceasta depinde de natura problemei ce trebuie soluţionată. Există două tipuri principale de probleme : probleme care implică o comparaţie între două sau mai multe moduri de organizare a aşteptării în anumite condiţii date şi probleme menite să determine estimaţii. Nu vom

insista asupra acestor aspecte, deoarece această deosebire între scopuri o vom analiza în cap. 8. Vom remarca totuşi că, din punctul de vedere al simulării, ele sînt foarte diferite. Pentru a înţelege mai exact de ce este aşa, vom considera problema distorsionării datelor de ieşire.

Simularea este, în esenţă, o operaţie de transfer. Ea necesită anumite date de intrare, cum ar fi repartiţia timpului de servire, tipurile de sosiri, regulile de determinare a disciplinei de aşteptare şi le converteşte în date de ieşire, ca de pildă repartiţia timpului total de aşteptare, proporţia clienţilor respinşi ş.a. Procesul real reprezentat de model constă de asemenea dintr-o operaţie de transfer ce acţionează asupra unor intrări similare calitativ, pentru a produce ieşiri similare. Datele de ieşire ale simulării pot diferi de cele reale în două sensuri : ele pot fi supuse erorilor sau pot fi supuse variabilităţii sau unei definiri incorecte. Altfel, spus, datele de ieşire provenite din operaţiile de simulare pot fi distorsionate. Ieşirile eronate sau defectuos definite sînt exemple de distorsiune. Există mai multe cauze ce generează aceste două tipuri de distorsiuni, cele mai importante fiind :

1. O simplificare exagerată sau pierderea în cadrul modelului a unor trăsături semnificative, care vor produce erori.

2. Modul greşit de generare a numerelor aleatoare.

Prima ar putea să nu rezulte din incompetenţa analistului de cercetări operaţionale, dar poate fi deliberată. Nu este întotdeauna necesar de a elimina erorile ce apar. Cea de-a doua poate fi controlabilă în anumite limite. Există o vastă literatură dedicată tehnicilor de reducere a dispersiei, ce nu pot fi prezentate aici.

Să considerăm un experiment de primul tip, experiment menit să compare două sisteme posibile de aşteptare. Să presupunem că problema constă în a compara eficienţa a două sisteme alternative de priorităţi de servire. Datele de intrare sînt în esenţă aceleaşi : tipurile de sosiri, repartiţiile timpului de servire, numărul staţiilor de servire. Datele de ieşire — sau, în terminologia utilizată anterior, măsura efectivităţii — sînt proporţiile clienţilor avînd de aşteptat mai mult de 1,2 respectiv 3 timpi medii de servire. (Cititorul îşi va reaminti că timpul mediu total de aşteptare nu este un parametru foarte indicat pentru asemenea situaţii.) Să presupunem că modelul de simulare produce date de ieşire distorsionate. Fie α_1, α_2 şi α_3 proporţiile reale în cele două cazuri, pentru primul sistem ; β_1, β_2 şi β_3

pentru sistemul al doilea. Ciclurile de simulare vor furniza următoarele rezultate :

Sistemul 1	Sistemul 2
$\alpha_1' + \eta_{11}$	$\beta_1' + \eta_{12}$
$\alpha_2' + \eta_{21}$	$\beta_2' + \eta_{22}$
$\alpha_3' + \eta_{31}$	$\beta_3' + \eta_{32}$

Valorile η reprezintă erorile statistic rezultate din generarea numerelor aleatoare, adică imprecizia sau variabilitatea. Ele au un fel de repartiţie ce poate fi estimată cu ajutorul repetării sau în alt mod. Problema care se pune este de a stabili dacă eroarea este aceeaşi pentru ambele serii ale ciclurilor, altfel spus, de a găsi cantităţile D_r şi R_r astfel încît

$$\alpha_r' - \alpha_r = D_r = \beta_r' - \beta_r$$

sau

$$\alpha_r'/\alpha_r = R_r = \beta_r'/\beta_r.$$

Dacă lucrurile se petrec astfel, existenţa erorilor nu are importanţă în compararea celor două sisteme, pentru că ele pot fi eliminate printr-o judicioasă alegere a metodei de comparare.

În cazul în care există erori proprii însăşi naturii alternativelor, va fi foarte dificil să le înlăturăm, sau chiar să le detectăm. În majoritatea modelelor de simulare a sistemelor de aşteptare ce au fost riguros construite, este foarte puţin probabil să apară erori corelate. Ca regulă generală, putem presupune că erorile sînt constante, fie aditive, fie multiplicative. Şi dacă acest lucru este adevărat, ele nu vor conta prea mult. Prezenţa lor nu va conduce la luarea unor decizii eronate.

În mod obişnuit comparaţiile trebuie făcute pentru diferite volume de trafic. Acest lucru ne conduce la proiectarea statistică a experimentelor. Pînă în prezent s-a realizat foarte puţin în proiectarea experimentelor de simulare şi au fost puţine cazurile descrise în literatură în care la experimentele de simulare s-a făcut în prealabil o proiectare a experimentului. Fără îndoială, că acesta este un domeniu promiţător pentru cercetările viitoare.

În opoziţie cu acele aplicaţii care reprezintă în primul rînd comparaţii între sisteme rivale, există alte aplicaţii în care se doreşte obţinerea unor estimări pentru anume cantităţi sau parametri. Ar fi important de ştiut, de exemplu, numărul angajaţilor ce vor fi utilizaţi în anul 1970 de compania Transcedental pentru rezervări de bilete, în fiecare oră din zi. Am văzut în cap. 1, că domnul Overlander a renunţat la serviciile acestei companii deoarece serviciul ei de rezervări telefonice nu a fost capabil să facă faţă traficului din orele de vîrf. Dacă presupunem că putem avea estimări ale volumului de trafic pentru 1970, problemă asupra căreia vom reveni în cadrul următorului capitol, atunci este esenţial faptul că modelul de aşteptare utilizat să convertească corect, fără erori, aceste estimări în număr de linii telefonice şi agenţi. Nu există o cale de înlăturare a erorilor prin realizarea unei comparări a naturii problemei : nu este vorba *de care* dintre ele, ci este vorba de *cît de multe.*

Simulările menite a răspunde la aceste tipuri de întrebări trebuie realizate cu foarte multă atenţie şi trebuie testate în condiţii cunoscute, sau pe modele matematice cunoscute, pentru a asigura eliminarea erorilor ce pot apărea în datele de ieşire. Aceasta este o problemă dificilă, ce a fost puţin analizată.

Problemele de cel de-al doilea tip creează dificultăţi şi în alt sens. În problemele de primul tip, chiar dacă aceasta s-a aplicat puţin, există cel puţin o metodologie de efectuare a experimentelor. Analiza dispersiei sau alte proceduri de decizie pot fi utilizate pentru a determina cea mai bună opţiune. Aici însă, în probleme de al doilea tip, ele nu pot fi utilizate.

Să presupunem că la un oficiu telefonic de rezervări de bilete avem un proces de aşteptare, pentru care $M/Lognormal/c$: (L/LIFO) este un model adecvat. Şi să mai presupunem că costul lunar de închiriere a unei linii este de A \$ şi cel de angajare a unui operator este de B \$, că există L linii şi c agenţi, astfel încît costul total lunar este

$$\mathcal{E} = AL + Bc.$$

Scopul este de a determina combinaţia pentru L şi c ce va conduce la minimizarea costului, asigurînd în acelaşi timp ca nu mai mult de $\alpha\,\%$ clienţi să fie respinşi datorită faptului că camera de aşteptare este plină, în timp ce dintre clienţii ce reuşesc să pătrundă în sistem cel mult $\beta\,\%$ au de aşteptat mai mult de V/μ timpi medii de servire.

Aceasta este problema determinării punctului (L_0, c_0) care minimizează pe $\mathscr{E}$ în regiunea plană definită de două relaţii generale de tipul $X(L, c) \leqslant \alpha$, $Y(L, c) \leqslant V/\mu$; dar forma explicită a funcţiilor $X(L, c)$ şi $Y(L, c)$ nu este cunoscută. S-ar părea că aceasta este o aplicaţie a unei tehnici iterative de căutare pe direcţii descrescătoare, dar este greu de realizat. Cititorului îi este recomandat să privească această problemă dintr-un punct propriu de vedere, după ce a analizat lucrările originale, legate de acest subiect, date la sfîrşitul acestei cărţi.

O metodă mai utilizată este de a crea graficul suprafeţelor X şi Y prin realizarea unor rulări de simulare pentru diferite combinaţii ale valorilor L şi c, şi apoi de a calcula costul $\mathscr{E}$ pentru fiecare pereche (L, c) care satisface restricţiile impuse lui X şi Y. În felul acesta, minimul poate fi determinat printr-o căutare şi nu una deosebit de dificilă. Invariabil, este necesar de a determina minimul prin interpolare.

Poate că cititorul va considera acest mod de lucru ca fiind costisitor şi imprecis. Dacă sînt create suficiente simulări pentru determinarea grafică perfectă a legăturilor dintre (L, c) şi (X, Y), astfel că interpolarea finală să determine valorile optime L_0 şi c_0, atunci este aproape sigur că costurile de calcul vor fi ridicate. Pe de altă parte, dacă se consideră un număr redus de rulări, precizia cu care se determină optimul (L_0, c_0) este foarte scăzută.

Oricum însă, această metodă este des utilizată şi a fost adoptată pentru studiul descris în cap. 15.

6.6. CONCLUZIE

Tehnicile de simulare sînt foarte utile atunci cînd se realizează comparaţii între forme alternative de organizare a facilităţilor de servire, iar modelele teoretice lipsesc. Aceasta ar putea fi chiar unica soluţie de rezolvare. Cu cît eroarea concluziilor unui proiect de aşteptare este mai mare, cu atît sînt mai utile tehnicile de simulare.

În celelalte cazuri, metodele de aproximare sînt mai clare, mai puţin costisitoare, mai rapide şi creează mai puţine dificultăţi.

7

NATURA SOLUȚIILOR OPERAȚIONALE

Atunci cînd un matematician afirmă că a rezolvat o problemă de așteptare, el înțelege prin aceasta că a construit un model al procesului real din care derivă problema, apoi a dedus să zicem o expresie formală a repartiției timpilor totali de așteptare în funcție de disciplina de așteptare și de repartiția sosirilor și timpilor de servire. Soluția poate lua forma unei ecuații integro-diferențiale, asemănătoare celor ale lui Takacs, ecuații care în general nu pot fi reduse la forme algebrice sau funcții cunoscute. Într-un anumit sens o asemenea expresie formală constituie soluția problemei; de aici se pot deduce anumite proprietăți ale modelelor de așteptare. Totuși aceasta nu este o soluție în sensul comun al cuvîntului, folosit de majoritatea oamenilor.

Practicienii din domeniul cercetării operaționale sînt preocupați de determinarea în timp util a soluțiilor pentru problemele ce apar în lumea reală. Fără îndoială că este foarte dificil să analizezi și să rezolvi astfel de sisteme, datorită numărului extrem de mare de parametri ce trebuie luați în considerare, a instabilității lor etc. De aceea, și în cercetarea operațională ca și în matematică, fizică sau economie, așa cum arătăm în cap. 2, este absolut necesar a se crea modele ideale. Aceste este primul pas spre formalizarea problemei; următorul pas este definirea conceptului de soluție. După cum în matematică Kronecker insista asupra faptului că teoremele de existență sînt inadmisibile, iar demonstrațiile matematice, pentru a fi acceptate, trebuie să fie constructive (adică, de exemplu, ele nu trebuie să arate doar că există o infinitate de numere prime, ci să indice și cum pot fi ele construite), tot așa, pentru a considera validă o soluție în lumea reală, ea trebuie să ne indice și modul de a acționa. Ideile pe care le

au practicienii cercetărilor operaționale despre natura soluțiilor problemelor pe care le încorporează în modele au un caracter special și sînt deduse din anumite concepții teoretice din fizica teoretică și științele economice.

Teoriile clasice ale fizicii și economiei se bazează pe două concepte : conceptul de echilibru și conceptul de optim. De exemplu, un sistem mecanic tinde spre o stare de echilibru, o stare în care o funcție obiectiv este minimizată. În teoria beneficiului, un patron acționează în funcție de modificările ce au loc pe piață, astfel încît să-și maximizeze beneficiul. Aceasta înseamnă că el caută o nouă stare de echilibru compatibil cu optimizarea unei funcții obiectiv. Aceste idei au pătruns și în cercetarea operațională. Morse și Kimball, în lucrarea lor „*Methods of Operational Research*", subliniază necesitatea găsirii unor măsuri ale eficacității sau a unor funcții obiectiv și a optimizării lor. În cercetarea operațională, soluție începe să însemne optimizarea unui model bine determinat.

Nemulțumirea analiștilor de cercetare operațională constă în aceea că în general executanții lucrărilor nu sînt foarte receptivi la indicațiile ce li se dau, ba uneori nici nu le dau prea mare atenție. Pentru a explica această stare de lucruri au fost formulate mai multe motive, s-a enunțat o întreagă gamă de opinii, de la autosubapreciere („nu știm cum să vindem produsele d-voastră") pînă la abuzivul „conducătorii și executanții sînt prea incapabili și reacționari". Cu toate că ar fi greșit să-i considerăm pe toți analiștii de cercetare operațională de talia lui Albert Laskers * sau că toți conducătorii sînt de talia lui Alfred Sloan**, nu acestea sînt cazurile unui fenomen atît de extins, aproape universal. Realitatea constă în existența unei divergențe de opinii dintre cercetători și executanți, în ceea ce privește înțelesul noțiunii de soluție.

În timp ce analiștii cercetării operaționale se plîngeau că executanții nu dau atenție sugestiilor lor, economiștii au avut nemulțumiri asemănătoare, că oamenii de afaceri nu le dau atenție.

În ultimii ani, mulți economiști, mai cu seamă din Statele Unite, au căutat să descopere adevăratele cauze ale neînțelegerilor iscate între ei și oamenii de afaceri. Cu toate că ei continuă să se plîngă că oamenii de afaceri ignoră teoria economică (ceea ce ar putea fi

---

* Un expert amerîcan, ce a realizat succese în domeniul reclamelor.
** Fostul patron al companiei General Motors.

adevărat), economiștii au încetat de a-i mai considera incapabili (ceea ce nu întotdeauna este adevărat). Economiștii au ajuns la două concluzii : prima este că tendința academică de a analiza problema, prin intermediul unor modele bine definite, nu este împărtășită de către oamenii de afaceri ; a doua, arată că oamenii de afaceri — chiar și antreprenorii — nu acționează și nu pot acționa consecvent conform unei singure reguli simpliste, așa cum este maximizarea beneficiilor.

Prima concluzie înseamnă următoarele : cei ce iau decizii în administrație, forțele armate, în afaceri și industrie nu gîndesc procesele operaționale și tendințele lor prin intermediul unor modele bine definite : imaginea lor este cea pe care o ai cînd te uiți prin sticlă mată.

A doua concluzie arată că executanții nu acționează ca pentru a atinge un punct final dorit ; cu alte cuvinte, să maximizeze funcția obiectiv așa cum este beneficiul. A fost lansată ideea că ei ar acționa de o asemenea manieră, încît să asigure cît mai mult posibil supraviețuirea întreprinderii. Baumol a mers mai departe cu această teorie, sesizînd că ei acționează ca pentru a maximiza beneficiul cu condiția satisfacerii anumitor restricții (de exemplu, rata de amortizare a investiției să nu fie sub 12 %). Aceasta este o idee excelentă, ce și-a găsit deja locul printre principiile și metodele cercetării operaționale, prin intermediul programării liniare. În spatele ei stă o impunătoare dezvoltare matematică, ce a început de la Lagrange.

Se mai poate spune că obiectivele oamenilor de afaceri sînt caleidoscopice. Ei au foarte multe scopuri, unele din ele chiar contradictorii, iar balanța importanței unuia sau altuia este în permanentă schimbare. Faptul că Baumol vorbește de o maximizare cu restricții implică existența mai multor factori în cadrul unei situații ce au tendința de a se opune unul altuia. Dacă toți cresc sau descresc împreună, nu ar mai fi necesară optimizarea prin introducerea unor restricții.

Și totuși, care este legătura dintre aceste lucruri, cercetarea operațională și soluția problemelor operaționale ?

Argumentele aduse de economiști indică motive pentru inacceptabilitatea multor rezultate ale cercetării operaționale, inclusiv aplicații ale teoriei așteptării. Și aceasta ca urmare în primul rînd a preocupărilor exagerate pe care teoreticienii le au pentru teorie și precizie atunci cînd prima este imposibilă, iar ultima inutilă și în al doilea rînd insistenței asupra găsirii unei funcții obiectiv care niciodată nu este în perfectă corespondență cu scopurile de compromis ale

executanților. Cu alte cuvinte, tehnica de obținere a soluțiilor prin procedeele clasice de optimizare nu este admisibilă în toate cazurile și aceasta nu pentru că este o tehnică șubredă, ci datorită variabilității caracteristicilor funcției obiectiv.

În scopuri practice modelele proceselor de așteptare sînt aplicate în trei direcții principale. :

1. În planificarea repartizării noilor facilități pentru operațiile în care așteptarea, sau aglomerarea, este un factor important.

2. În evaluarea strategiilor de desfășurare a forței de muncă pentru facilitățile existente, astfel încît să se realizeze condiții de servire standard și cu preț redus.

3. Pentru compararea diferitelor forme de organizare a facilităților de servire.

Este neîndoielnic că problemele tehnice ale formării și analizei modelelor de așteptare, menite să previzioneze necesarul de capacitate a noilor facilități, sînt foarte anevoioase. În cap. 13 vom analiza o asemenea situație. Dar modelele de așteptare nu sînt singura sursă de dificultăți. O mare importanță trebuie acordată previziunii parametrilor modelelor. Sigur că prognoza poate fi uneori hazardată, și este probabil să se utilizeze una ce nu dă rezultate satisfăcătoare. De exemplu, chiar și cu ajutorul unor tehnici cum ar fi ajustarea exponențială a seriilor de timp (dar nu încă a modelelor econometrice) se consideră a fi satisfăcătoare o prognoză pentru un an, ce dă traficul lunar de curse aeriene de pasageri dintr-un aeroport, cu o abatere de 5% (în plus sau în minus).

Prognozele orare au un grad și mai mare de nesiguranță. Remarcați cît de dificil este a determina, să spunem cu cinci ani înainte, costul planificării și realizării unui nou aeroport. O eroare de 25% în prognozele zborurilor lunare ar putea fi, în acest caz, satisfăcătoare. În plus, ar fi dificil de prevăzut exact care va fi timpul mediu de servire, mai cu seamă dacă activitățile ce constituie operația de servire sînt noi, sau aranjate într-o formă nouă. (În cap. 10 a fost ales un exemplu de acest fel.) Chiar și cel mai competent inginer nu este dotat cu puteri supranaturale!

În aceste condiții căutarea unui model de așteptare elegant și riguros este greșită.

Inevitabil, la sfîrșitul lungilor și dificilelor serii de calcule ale cerințelor viitoare, în care precizia depinde direct de volumul de date de previzionat, de durata de studiu și de conștiință profesională,

inginerii consultanți sau arhitecții, bazîndu-se pe experiența lor, vor adăuga cel puțin 10% la fiecare calcul. Cu alte cuvinte, din punct de vedere practic, este foarte neeconomic să se determine conștiincios și exact valorile strategice ale modelului, în condițiile unei planificări pe termen mai lung. Este însă esențial să dăm răspuns la întrebarea : care trebuie să fie gradul de precizie al modelului, sau cît putem noi realiza?

A doua clasă de utilizare a modelelor de așteptare constă din aplicațiile menite să determine desfășurarea facilităților existente. Se remarcă că aceasta reprezintă o situație cu totul diferită de cea prezentată anterior. Ne putem imagina în acest caz că stațiile de servire există deja și problema care se pune este de a determina numărul celor ce lucrează în diferite momente ale zilei, în diferite zile ale săptămînii etc., pentru a crea un serviciu standard.

Acest lucru poate fi realizat cu ajutorul prognozei pe termen scurt a volumului de trafic (utilizînd traficul în sens Erlang) pusă la îndemîna managerului.

Trebuie totuși să menționăm că de obicei un manager preferă să utilizeze formule foarte simple care să-l ghideze printre multiplele sarcini care le are de îndeplinit. În acest caz, simplitatea înseamnă aproximare. În cap. 15 se va analiza un astfel de caz. Un model cu un înalt grad de fidelitate în reprezentarea realității va fi foarte complex, cu grad mare de precizie și deci aproape inutil. Un model cu un grad scăzut de fidelitate în reprezentarea realității poate fi mai simplu, nu tot atît de precis, dar util. Cititorul va decide care ar trebui să fie alegerea făcută de către un analist de cercetare operațională.

În sfîrșit, au rămas problemele de comportare a diferitelor forme de organizare a facilităților de servire. Să presupunem că există trei posibilități de combinare. (Cititorul este rugat să-și amintească sugestiile d-lui Murdo Macdonald din cap. 1 sau să se uite peste exemplul dat în cap. 9.) Să presupunem în continuare că traficul, exprimat în erlangi, este constant și independent de natura combinației. Pentru simplificare, să considerăm că problema este de a decide care variantă produce cel mai scăzut timp mediu de așteptare, pentru un număr dat de stații de servire.

Cîteodată este posibil să se răspundă la această întrebare fără a face nici un calcul, bazîndu-ne doar pe experiență. De exemplu, presupunînd că repartiția timpului de servire nu este afectată de natura variantei alese, atunci este limpede că o variantă ce va crea

o utilizare deplină a stațiilor va fi preferată altei combinații ce nu implică acest lucru. Sau o variantă care încurajează mobilitatea clienților între șirurile de tipul „bilet-ghișeu" va fi preferată unei variante ce nu realizează acest lucru, deși nu este atît de favorabilă ca un sistem „maître d'hôtel", cu posibilitatea maximă de servire. În asemenea condiții, sînt satisfăcătoare și concluziile generale rezultate chiar din modelele deduse din ecuații integro-diferențiale. Problema ce se pune este de a identifica repartiția timpului de servire, ordinea sosirilor și disciplina de așteptare și nu de a le calcula precis.

În alte condiții nu este suficient să cunoști care combinație este mai eficientă; trebuie cunoscut și cu cît este mai eficientă o variantă decît celelalte, căci costurile asociate variantelor pot diferi. Aceasta introduce un nou element al problemei, care poate fi reprezentat în felul următor:

Variantă	Costul forței de muncă/om	Cheltuieli de întreținere/an	Timp mediu de așteptare/ client
1	45	20	$1,1/\mu$
2	15	40	$1,2/\mu$
3	10	50	$0,9/\mu$

Termenul „cheltuieli de întreținere" semnifică cheltuielile de achiziție a utilajului (instalației) și de menținere în funcțiune. Cheltuielile de întreținere includ deprecierea și cheltuielile de menținere care, pentru orice combinație, se presupun a fi constante. Tabelul arată comparațiile realizate pentru timpi de așteptare egali. Variația cifrelor din ultima coloană este consecința faptului că numărul de angajați este o variabilă discretă. De aceea, este practic imposibil a egaliza timpii medii de așteptare prin manevrarea forței de muncă. Remarcăm că, din punct de vedere financiar, a doua variantă este cea mai favorabilă. Aceasta este o situație clasică. Totuși costul forței de muncă poate varia ca rezultat al unor negocieri asupra salariului, în timp ce cheltuielile de întreținere rămîn aceleași. În acest caz problema este: în ce condiții de modificări ale costurilor, varianta a doua rămîne cea mai bună? Cititorul va remarca faptul că varianta secundă continuă a fi cea favorabilă pînă cînd costul forței de muncă se dublează. Probabil că aceasta nu se produce curent. și dacă este așa, putem spune că „soluția" — potrivit căreia varianta secundă este cea mai bună — nu este influențată de costuri. Pot apărea însă și alte efecte.

De exemplu, rata creșterii forței de muncă cu volumul traficului, pentru menținerea unui serviciu standard, poate diferi de la o variantă la alta. Acesta este un fenomen des întîlnit, în special cînd timpii medii de servire diferă între variante (vezi cap. 9). Astfel, o creștere de 15% a volumului traficului poate da următoarele rezultate :

Varianta	Costul forței de muncă/an	Cheltuielile patronului/an	Timpul mediu de așteptare/client
1	60	25	1,0/μ
2	25	43	0,8/μ
3	13	52	0,9/μ

Varianta a treia este acum cea mai eficientă. „Soluția" pe care am găsit-o este influențată de modificările volumului traficului.

Exemplele prezentate sînt artificiale ; dar ele se bazează pe o experiență suficient de convingătoare. Cazurile ce vor fi prezentate în capitolele ce urmează vor ilustra fenomenele posibile. Cînd se constată că previziunile pot fi eronate, că restricțiile operaționale se pot modifica și că orice presupunere făcută asupra valorilor parametrilor poate fi invalidată de evenimente, este lesne de înțeles că „soluția" ce trebuie căutată este caracterizată prin robustețe.

Punctul de vedere ce a stat la baza acestei argumentări este următorul : este puțin recomandabil a construi modelul unui proces și prin analizarea sa să se tragă concluzii definitive — ca de exemplu, „portul necesită două noi dane pînă în 1970" — fără a verifica senzitivitatea modelului. Unele modificări făcute valorilor parametrilor au un efect important asupra intensităților traficului. În aceste condiții, nu există distincții clare ce trebuie făcute între unele răspunsuri și altele. Astfel și-au făcut apariția mulțimile vagi (fuzzy) ale bussinesmanului.

De exemplu, dacă curbele caracteristice costului (cheltuielile forței de muncă plus cheltuielile de întreținere) ar fi fost proiectate pentru cele trei sisteme diferite din exemplul nostru în condițiile unui trafic diferit, pentru fiecare s-ar fi găsit un alt punct de minim. Dar, deoarece nu se poate cunoaște exact dacă coeficienții funcției costurilor au fost corect determinați, pot apărea anumite oscilații verticale. Pentru diferite valori ale acestor coeficienți, curbele se vor intersecta în puncte diferite, iar minimul se va modifica, ca în exemplul considerat. Astfel, sînt introduse și oscilații orizontale. În aceste

condiții, este puțin probabil ca modelele foarte sofisticate să servească scopului ce ni-l propunem.

O soluție operațională a unei probleme de așteptare este orice soluție suficient de judicios evaluată pentru a conduce la o decizie corectă. Ea poate să nu se bazeze pe un model foarte complicat. Ea poate fi dedusă și din anumite constatări evident neadevărate legate de natura repartiției timpului de servire. Lucrul cel mai important în cercetarea operațională nu este de a obține soluții în sens matematic, ci soluții operaționale, ce vor conduce la elaborarea unor decizii valide, executabile.

8

TREI TIPURI DE APROXIMĂRI

În cadrul capitolelor precedente am prezentat unele probleme teoretice ; cele ce urmează au legătură directă cu aplicaţiile practice. Acest scurt capitol este o punte de legătură între ele şi este un prilej de a relua argumentarea fundamentală a acestei lucrări.

În principiu, atît din punct de vedere teoretic cît şi estetic, varianta cea mai dorită de rezolvare a problemelor legate de sistemele de aşteptare constă în a construi un model matematic foarte fidel şi a determina soluţii prin analiza acestui model. S-a spus suficient de mult despre dificultăţile matematice ce intervin în analiza modelelor, chiar şi pentru cele mai simple sisteme, astfel încît se pune problema : „Tratarea ideală este oare posibilă?". Examinarea analizelor matematice făcute în capitolele 2 şi 5 pare a duce la concluzia că numai modelele elementare pot fi supuse unei analize matematice exacte şi că aceste modele nu au legătură cu procesele ce se întîlnesc frecvent în lumea reală. Cazurile ce vor fi prezentate în capitolele următoare vor demonstra că sistemele cotidiene de aşteptare nu au o structură atît de complicată ca cea a modelelor de aşteptare ce au fost amplu analizate pînă în prezent. Putem conchide că tratarea ideală, tradiţionala procedură ultraştiinţifică a cercetării operaţionale, în general, nu poate fi realizată. Se pare că reprezentarea sistemelor de aşteptare cu ajutorul modelelor matematice foarte fidele nu poate fi realizată adesea, iar dacă da, atunci modelele nu pot fi analizate profund. Aceasta a dus la crearea tehnicilor de simulare. Am vorbit în cap. 6 despre aceste tehnici pentru a sugera că, folosindu-le, nu se pot obţine chiar atît de uşor pe cît se pare rezultate exacte privind sistemele de aşteptare. Apar mai multe dificultăţi : comportarea tranzitorie, mărimea selecţiilor (lucrarea lui Gebhart, prezentată în

anexa 1, este recomandată celor care nu cred că aceasta ar reprezenta o dificultate) și proiectarea experimentală.

Atît procedura tradițională cît și tehnica simulării tind spre a atinge precizia maximă. Procedura tradițională nu este adesea utilizabilă; simularea însă da, și se pot obține răspunsuri suficient de precise, dar cu un cost de calcul relativ ridicat. Din păcate, tocmai în astfel de situații, care au o mare importanță practică — ca de pildă, cele legate de intensitățile mari ale traficului — ciclurile de simulare trebuie să fie foarte lungi, pentru a atinge performanțele dorite. În capitolul precedent ne întrebăm „Oare este atît de necesară o precizie ridicată?". Și răspunsul la această întrebare era că în multe situații nu este necesară, iar pentru a indica de ce trebuie să fie așa a fost introdusă ideea de soluție operațională. Conceptul de soluție operațională nu este analizat *in extenso* în această lucrare: cititorul va fi capabil să o raporteze la cercetările întreprinse în domeniul proceselor de decizie executivă și al teoriei deciziei.

Odată acceptată ideea că a căuta modele exacte și riguroase, ce vor determina rezultate numerice sau analitice neperturbate, este o greșeală, și că o aproximare cu un grad acceptabil de precizie este calea de urmat, atunci putem privi în cu totul alt mod atît procedura de construire a modelului matematic cît și tehnicile de simulare.

Argumentul acestei lucrări este că teoria așteptării este utilă. În pofida complexității matematice, pe de o parte, și unor inadvertențe practice, pe de altă parte, teoria în forma ei actuală poate determina soluții operaționale pentru multe situații. Aceasta cere însă ca să se accepte necesitatea aproximațiilor de primul tip, adică aproximarea prin utilizarea modelelor foarte simplificate. Alt mod de redare a acestei idei este că primul tip de aproximare constă în a obține soluții operaționale prin apelarea modelelor care sînt, apriori, eronate, dar sînt suficient de simple pentru a le putea manipula. Capitolele ce urmează, și în special capitolele 10, 11 și 12, conțin cazuri practice ce vor arăta aplicabilitatea acestei proceduri. Nu o să mai insistăm acum asupra acestor lucruri, exceptînd faptul că pentru un analist de cercetare operațională procedura ar fi deosebit de utilă, poate cea mai utilă dintre toate.

În cazul în care aproximarea de primul tip este considerată a fi inacceptabilă, este necesar a utiliza simularea. Aceasta ar fi necesară în următoarele condiții: în cazul că riscul luării unei decizii eronate datorită modelului foarte simplificat nu este cunoscut, sau în cazul că penalizările pentru o decizie eronată sînt foarte mari. Prima dintre

aceste condiţii ţine de experienţa cercetătorului : în cercetarea operaţională continuă să aibă o pondere importantă arta şi îndemînarea cu care lucrezi.

A doua este un criteriu mai obiectiv. Modelele de simulare sînt o formă a aproximării, dar aproximarea este de un tip cu totul diferit aceleia rezultate din procedurile clasice de construire a modelelor. Utilizînd un mare volum de muncă, este posibil a simula procesul de aşteptare de o manieră foarte apropiată de realitate. Fidelitatea modelului de simulare poate fi mai ridicat decît a oricărui alt model construit prin metode matematice pure. Calitatea aproximativă a simulării este o consecinţă, aşa cum am văzut în capitolul precedent, a generării de numere aleatoare, adică a introducerii erorilor statistice de selecţie. Dar, prin multiple încercări, acestea pot fi reduse şi în principiu simulările pot avea un grad de precizie oricît de înalt. Din păcate, nu este întotdeauna posibil a aloca sume imense corespunzătoare costului unui număr foarte mare de cicluri de simulare şi de aceea este recomandabil să se utilizeze al doilea tip de aproximaţie. Şi deoarece pentru a găsi o soluţie operaţională nu este întotdeauna absolut necesar un rafinament extrem, aproximarea de tipul al doilea este validă şi folositoare.

Principiul aproximării de tipul al doilea este următorul : un sistem este simulat în anumite condiţii, astfel încît se obţin puncte pe suprafeţele de răspuns ce interesează. Acestea devin apoi obiectul analizelor statistice cu scopul determinării unor relaţii empirice dintre diferiţii factori (staţii de servire, traficul, priorităţile, timpii de servire etc.), ce pot fi apoi trataţi ca un model de ordinul doi al sistemului. Este o înlocuire a experimentării sistemelor din lumea reală. Un exemplu de utilizare al acestei tehnici este prezentat în cap. 15.

Al treilea tip de aproximare este cel clasic, utilizat în matematică, fizică şi în mai toate domeniile. Acesta este exemplificat prin două metode :

1. Dezvoltarea unor funcţii imposibil de manipulat în serii de puteri, în care termenii de rang mare pot fi ignoraţi, pentru anumite domenii de variaţie ale variabilelor.

2. Substituirea valorilor iniţiale în formulă, urmată de procesul iterativ.

Aceste proceduri clasice au fost utilizate în studiul modelelor de aşteptare, dar nu atît de mult cum s-ar putea crede. Avantajul acestui tip de aproximare (comaparativ cu celelalte două) este relativa uşurinţă în estimarea erorilor rezultatelor. Probabil că cea mai specta-

culoasă utilizare a acestui tip de aproximare a fost realizată de Riordan, care a ajuns la formulele, prezentate în capitolele 3 și 4, pentru repartițiile timpului de așteptare în cazuî modelelor $M/M/1$: (∞/SIRO) și $M/M/c$: (∞/SIRO), formule care au constituit baza calculelor făcute pentru realizarea graficelor necesare determinării rapide a rezultatelor acestor modele.

Merită să mai menționăm alte două exemple de utilizare al primului tip de aproximare matematică și unul pentru al doilea tip; toate trei se referă la modelul de așteptare $M/D/c$: (∞/FIFO). Molina a determinat următoarele formule aproximative pentru timpul mediu de așteptare și pentru probabilitatea ca toate stațiile de servire să fie ocupate :

$$W_q = P(t > 0)\,\frac{1}{\mu c(1-\rho)}\,\frac{c}{c+1}\,\frac{1-\rho^{c-1}}{1-\rho^c},$$

$$P(t > 0) = \frac{\dfrac{\lambda^c e^{-\lambda}}{c!}\,\dfrac{c}{c-\lambda}}{1 - e^{-\lambda}\sum_{n=c}^{\infty}(\lambda^n/n!) + \dfrac{\lambda^c e^{-\lambda}}{c!}\cdot\dfrac{c}{c-\lambda}}. \qquad (8.1)$$

Pollaczek a determinat formule mai simple pentru aceleași mărimi ce sînt folositoare în special cînd există un număr mare de stații de servire. Rezultatele obținute sînt :

$$W_q = \frac{1}{c\sqrt{2\pi c}}\,\frac{\rho^c e^{(1-\rho)c}}{(1-\rho)^2}\left[1 + o\left(\frac{1}{c}\right)\right],$$

$$P(t > 0) = 1 - W_q(0) = \frac{1}{1-\rho}\,\frac{\rho^c e^{(1-\rho)c}}{\rho\sqrt{2\pi c}}. \qquad (8.2)$$

Cititorul va remarca că utilizînd aproximarea factorialelor dată de Stirling este posibil a ajunge la rezultatul lui Pollaczek pentru W_q plecînd de la rezultatul lui Molina.

În sfîrșit, în 1953 Everett publică o procedură iterativă de calcul a probabilităților stării stabile, p_n, pentru acest model de așteptare.

Procedura sa, care este prea vastă pentru a o prezenta complet, depinde de observaţia că pentru valori mari ale lui n

$$p_{n+c} = \psi^c p_n.$$

Această relaţie a fost combinată cu ecuaţiile lui Crommelin pentru p_n, deducîndu-se o relaţie de determinare a lui ψ, ce pare a juca un rol în modelul $M/D/c : (\infty/\mathrm{GD})$ similar cu cel jucat de ρ în modelul $M/M/c : (\infty/\mathrm{GD})$. Ca urmare, Everett a introdus valoarea sa pentru ψ în formulele lui Erlang legate de modelul anterior, pentru a obţine noi estimaţii pentru p_n. Aceste valori sînt apoi utilizate în ecuaţiile modelului cu timpi constanţi de servire, procesul fiind iterat pînă la atingerea unor estimaţii stabile. Apare astfel necesitatea ca cercetătorii să dezvolte aproximaţiile matematice de acest tip şi este surprinzător că pînă acum nu s-a acordat mai multă atenţie acestor probleme.

Cu aceasta am încheiat partea teoretică şi „filozofică" a lucrării. Următoarele capitole, exceptînd concluziile, tratează exclusiv probleme ale aplicaţiilor practice.

9

SISTEME DE REZERVARE DE LOCURI

Primul dintre sistemele reale de aşteptare pe care le vom analiza este cel observat pe parcursul studiului sistemului de transport de mărfuri al unei companii aeriene *. Fără îndoială că există sisteme similare şi în alte situaţii, iar problemele legate de acest sistem sînt, probabil, printre cele mai comune pe care le întîlneşte un analist de cercetare operaţională. Sistemul în sine este destul de adecvat pentru aplicarea teoriei aşteptării, astfel încît este util să începem analiza cu o astfel de problemă.

9.1. DESCRIEREA PROBLEMEI

Într-unul din oraşele principale pe care le serveşte, Compania aviatică Europa Airways Corporation (EAC) întreţine un vast departament pentru organizarea transportului mărfurilor. Atunci cînd un expeditor doreşte să obţină spaţiu în cadrul unui zbor EAC, spre un alt oraş, îl poate obţine chemînd la telefon Centrul de rezervări EAC, existent în cadrul aeroportului (detaliile privind procedura de rezer-

---

* Acest capitol a fost mult simplificat de două studii realizate de autor şi de colegii de la BEA în 1957 şi 1958. S-a utilizat o formă narativă pentru a accentua faptul că evenimentele şi cauzele lor au fost rearanjate şi prezentate într-o altă formă. S-au menţinut reale datele şi legăturile existente. Materialul a fost prezentat cu acordul BEA.

vare nu ne interesează). Toate apelurile provenite de la clienţi sînt direcţionate, prin intermediul unui tablou de distribuţie, spre un grup de zece linii telefonice, iar prin intermediul acestor linii apelul poate ajunge la cinci angajaţi. Adică oricare dintre cei cinci angajaţi pot rezolva un apel sosit pe oricare dintre cele zece linii. Dacă apelul soseşte atunci cînd tabloul de distribuţie este blocat, apelul va fi respins sau pierdut; dar acest caz are loc suficient de rar şi-l vom putea ignora. Dacă toate cele zece linii sînt ocupate, apelul este reţinut la tabloul de distribuţie; dacă există linii libere, dar toţi agenţii sînt ocupaţi, apelul este reţinut pe una dintre liniile libere. În ambele cazuri apelul aşteaptă.

Departamentul de relaţii cu publicul din cadrul EAC a primit o serie de plîngeri din partea clienţilor că, într-o anumită perioadă a zilei, ei trebuie să aştepte excesiv de mult înainte ca apelurile lor să fie preluate de către agenţi. În acelaşi timp în cadrul departamentului au apărut anumite probleme de utilizare a forţei de muncă, care implică necesitatea unei reorganizări radicale. Ca urmare, directorul general a elaborat un proiect de reorganizare care, printre altele, afectează şi serviciul de rezervări telefonice. În proiectul său se arată că activitatea serviciului de rezervare poate fi clasificată în funcţie de destinaţia geografică a coletelor şi, după o scurtă analiză, el a stabilit cinci categorii de trimiteri, anume spre: 1) Insulele britanice; 2) nordul şi estul Europei; 3) sudul Europei; 4) Orientul Apropiat; 5) Africa. În proiectul său va exista cîte o secţie pentru fiecare dintre aceste categorii de trimiteri, iar în cadrul fiecărei secţii angajaţii vor efectua rezervări, înregistrări, întocmiri de facturi etc. Centrul de rezervări actual va fi desfiinţat şi fiecare secţie va avea propriile linii telefonice.

În timpul discuţiilor purtate directorul general al Companiei EAC a relevat faptul că traficul telefonic legat de fiecare dintre cele cinci zone este diferit (de exemplu există de două ori mai multe apeluri pentru trimiteri de colete către nordul şi estul Europei, decît cele spre Orientul Apropiat) şi că existenţa unui singur telefon în fiecare secţie nu va asigura un serviciu egal tuturor clienţilor. El susţine că probabil anumite secţii vor necesita două sau chiar trei linii telefonice şi angajaţi. Astfel, numărul total de angajaţi în cele cinci secţii poate fi mai mare decît cinci. Se pune întrebarea: „Poate să-şi permită compania o asemenea creştere de personal"? În acest moment putem considera că este cerută părerea unui analist de cercetare operaţională (pe care-l vom numi dr. Wiley).

9.2. PRIMA FAZĂ

Primul pas făcut de dr. Wiley a fost de a vizita Centrul de rezervări în scopul de a observa operațiile curente. El a constatat astfel că în fața fiecărui angajat există un sistem de vizualizare în care fiecare dintre cele zece linii este reprezentată de un bec alb și unul roșu. Atunci cînd un apel așteaptă pe una dintre linii, se aprinde becul roșu corespunzător; iar cînd apelul a fost preluat, se aprinde becul alb. În consecință, este destul de ușor să culegem date privind timpii totali de așteptare la telefon; timpul total de așteptare al unui apel se poate cronometra între momentul aprinderii becului roșu și momentul aprinderii becului alb corespunzător.

La apeluri nu se răspunde în ordine, căci angajaților le este greu să rețină ordinea de așteptare. Apelurile sînt rezolvate, aparent, în ordine aleatoare.

Dr. Wiley a decis că trebuie să aibă anumite date înainte de a continua analiza; din păcate însă liniile telefonice nu au dispozitive de măsurare; cu toate acestea, el găsește posibilitatea de a aranja cu șeful centralei telefonice a aeroportului ținerea evidenței tuturor apelurilor către centrul de rezervări în timp de o săptămînă. La sfîr-

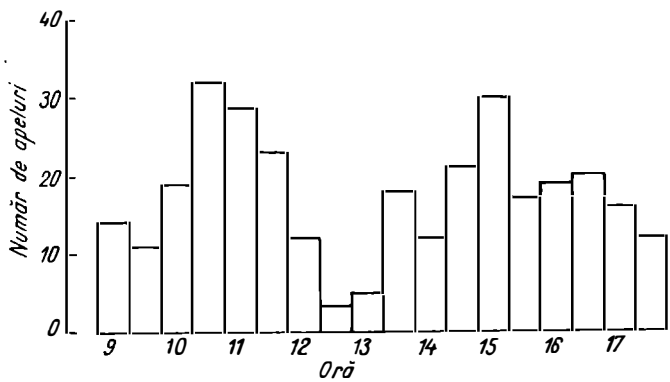


Fig. 9.1. Apeluri înregistrate la centrul de rezervare (8 martie 1957).

șitul săptămînii primește datele și constată că numărul apelurilor în zilele de luni pînă vineri este repartizat în același mod (v. fig. 9.1). Sîmbătă și duminică apelurile sînt mai puțin numeroase iar repartiția

lor este alta. Dr. Wiley şi-a dat seama că nu sfîrşitul de săptămînă trebuie avut în vedere, deoarece atunci sînt probleme mai puţin numeroase. Trebuie acordată atenţie doar celor cinci zile lucrătoare; în special trebuie concentrată atenţia asupra comportamentului sistemului în *perioadele de aglomeraţie*, pentru că în aceste perioade este cel mai probabil ca sistemul să fie supraîncărcat. Diagrama arată că perioadele ocupate din zi sînt între 10,30—12 şi 14,30—15,30.

Următorul pas este de a observa sistemul în timpul perioadelor de aglomeraţie. Dr. Wiley şi asistentul său reuşesc să înregistreze următoarele informaţii legate de apelurile din perioadele de aglomeraţie, folosind sistemul de vizualizare al angajaţilor:

intervalele dintre apelurile consecutive;

timpii de aşteptare pentru mai multe apeluri;

timpii de servire pentru mai multe apeluri.

Intervalele dintre apelurile consecutive, pe parcursul a patru intervale de jumătate de oră din perioadele ocupate, sînt testate prin testul indicelui de dispersie Poisson bazat pe gruparea în 15 intervale de cîte 2 minute. Toate cele patru eşantioane ale apelurilor dau dreptul să se facă ipoteza că sosirile sînt repartizate Poisson.

Există două tipuri de apeluri telefonice: 10% sînt informative şi sînt rezolvate foarte rapid (de exemplu: „Mîine există vreo cursă spre Praga?"), celelalte sînt apeluri pentru rezervări de spaţii de transport.

Repartiţiile timpilor de servire pentru ambele tipuri de apeluri sînt unimodale în apropierea lui zero şi, după cum a putut remarca cu uşurinţă Wiley, pot fi suficient de bine aproximate cu repartiţii exponenţiale, cu mediile de 3,2 şi respectiv 3,7 minute.

Dr. Wiley a remarcat de asemenea un fapt semnificativ în legătură cu activitatea depusă de un angajat atunci cînd răspunde la un apel de rezervare. În aproape toate cazurile, agentul trebuie să consulte unul din fişierele, dealtfel excelent proiectate, în care să verifice dacă există spaţiu disponibil la zborurile pentru care este interesat clientul său. Un desen al poziţiei acestor fişiere şi al angajaţilor este prezentat în fig. 9.2. Cititorul va remarca, aşa cum a făcut-o şi dr. Wiley, că în mod frecvent agenţii trebuie să-şi părăsească locurile pentru a consulta fişierele şi aceasta este o componentă importantă a timpului de servire.

În acest moment analistul nostru se retrage în biroul său pentru a analiza cele observate. Este clar că sistemul existent poate fi clasi-

ficat ca fiind de tipul $M/M/5 : (\infty/\text{SIRO})$; deci un model care nu creează probleme deosebite şi care a fost foarte mult studiat*. Care este sistemul propus? În loc de un grup de zece linii cu cinci angajaţi, disponibili oricînd, vor fi cinci secţii, fiecare de tipul $M/M/c : (L/?)$; din punctul de vedere al analizei, un rol important îl joacă c, L şi

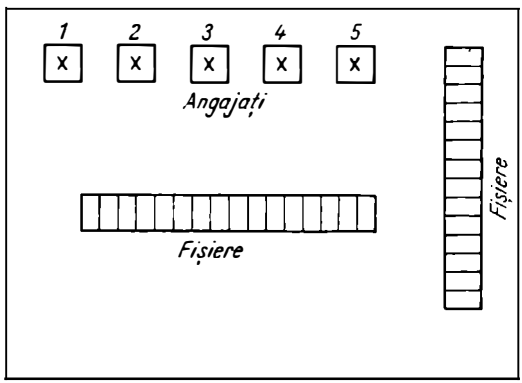


Fig. 9.2. Schiţa oficiului de rezervare.

regula încă nedeterminată a disciplinei de aşteptare. Nu se poate încă stabili nimic precis. Dr. Wiley mai face cîteva verificări asupra sistemului existent, utilizînd datele culese. Tabelul 9.1 prezintă cîteva date legate de timpii de aşteptare, înregistrate pe parcursul celor patru intervale de jumătate de oră în timpul perioadelor de aglomeraţie, atunci cînd frecvenţa apelurilor este de aproximativ unul pe minut. (Nu toate datele culese pot fi utilizate, deoarece frecvenţele medii de sosire sînt prea diferite între ele pentru a putea fi comparate.) Prima coloană a tabelului prezintă rezultatele unui model teoretic de tipul $M/M/5 : (\infty/\text{SIRO})$. Corespondenţa este suficient de corectă şi ne putem imagina satisfacţia dr-lui Wiley în faţa obţinerii unei aproximări satisfăcătoare a realităţii.

* Capacitatea sistemului poate fi considerată infinită şi nu doar 10, deoarece apelurile pot fi menţinute în tabloul de distribuire. Dacă $L = 10$, ar putea să apară unele dificultăţi.

Tabelul 9.1

Timpi observaţi şi teoretici de aşteptare

M/M/5 ; (∞/SIRO)		Intervalul 1	Intervalul 2	Intervalul 3	Intervalul 4
W_q	1,30	1,10	1,80	1,43	1,35
			rezultate comasate		
$P(v > 5$ min)	0,17		0,13		

Analistul nostru consideră că acum poate să se gîndească la noua propunere. Dr. Wiley începe prin a culege mai multe date. (Atît conducătorii cît şi personalul de execuţie au întotdeauna o mare nedumerire asupra volumului mare de informaţii numerice, pe care îl necesită cercetarea operaţională.) Făcînd observaţii asupra liniilor telefonice, pe baza unui plan de culegere eşantionată a acestora, pe o perioadă de două săptămîni, el obţine următoarele rezultate privind apelurile telefonice relative la zona geografică a destinaţiei transportului solicitat :

1. Insulele britanice 17,0 %
2. Nordul şi estul Europei 21,6 %
3. Sudul Europei 28,6 %
4. Orientul Apropiat 12,2 %
5. Africa 26,6 %

Zona 3 primeşte mai mult decît dublul comenzilor pentru zona 4. Dr. Wiley verifică apoi dacă acest aspect al repartizării procentuale a apelurilor pe zone de destinaţie este în concordanţă cu transporturile efectuate şi găseşte că este. Mai mult, el constată că repartizarea cererilor de transport spre diferite zone de destinaţie este aproximativ constantă în timpul perioadei de vîrf a verii şi de asemenea în timpul perioadei de vîrf premergătoare anului nou. Aceasta îi întăreşte convingerea că procentele determinate pentru numărul de apeluri aferente diferitelor zone de destinaţie sînt aplicabile şi perioadelor de vîrf ale anului. (Nu e nici o îndoială că această procedură are la bază presupuneri foarte „tari", dar ce altceva poată să facă?)

Dr. Wiley mai face încă o verificare, căci este posibil ca procentele determinate să fie corecte în sensul mediei, dar numărul de apeluri pentru diferitele zone de aşteptare pot să atingă maximul la ore diferite în timpul zilei, astfel încît, la un moment dat al zilei, pot fi mai multe apeluri pentru zona 4, decît pentru zona 3. Aceasta nu ar fi avut importanţă în vechiul sistem, dar va fi foarte important dacă

va fi adoptat noul sistem. Analiza înregistrărilor făcute de el însuşi şi de asistentul său îi arată că deşi sînt oarecari neconcordanţe a momentelor de vîrf, rapartizarea procentuală obţinută pentru apelurile către diferite zone, relativă la perioadele de vîrf ale primelor cinci zile ale săptămînii, este în esenţă corectă. Dr. Wiley va putea face o serie de calcule numai dacă rezolvă problema disciplinei de aşteptare ce va exista în noul sistem. Directorul companiei ce răspunde de telecomunicaţii îi spune să se prevadă înlocuirea tabloului central de distribuţie cu unul automat. Un echipament automat va funcţiona pentru fiecare linie, şi nici un apel nu va mai putea fi menţinut în tabloul de distribuţie. Deci regula FIFO va determina disciplina de aşteptare. Dacă va fi menţinut centrul de rezervări, adică sistemul actual, el va dispune de o capacitate de 20, avînd cel mult zece linii telefonice. Dacă va fi adoptat noul sistem, atunci fiecare secţie va dispune de cinci. Dr. Wiley ştie că modelul cel mai apropiat pentru noul sistem ce a fost propus este, pentru fiecare din cele cinci zone, $M/M/c$: (5/FIFO). La un nivel al traficului de 3,7 erlangi, intrările în modelul $M/M/5$: (20/FIFO) reprezentînd sistemul existent sînt de 3,7 erlangi, iar intrările pentru fiecare zonă propusă sînt respectiv 0,64 ; 0,80 ; 1,06 ; 0,45 şi 0,75 erlangi, corespunzînd procentelor de apeluri ce au fost stabilite. Se pot face comparaţii la acest nivel al traficului, deoarece acesta este o reprezentare rezonabilă a încărcării ce trebuie rezolvată ; putem compara timpii medii de aşteptare, probabilităţile de pierdere şi probabilităţile ca un timp de aşteptare să depăşească un minut. Scopul constă în a calcua aceste cantităţi pentru sistemul actual cu cinci staţii de servire şi apoi de a determina necesarul de staţii de servire pentru diferitele zone ale sistemului propus, capabil să asigure aceeaşi calitate a serviciului.

Pentru modelul $M/M/c$: (10/SIRO) nu există o formulă convenabilă, astfel că Wiley va trebui să obţină rezultate prin simulare. Pentru modelul $M/M/5$: (20/FIFO) ca şi pentru $M/M/c$: (5/FIFO) există formula (4.6). Cu aceste mijloace el a obţinut rezultatele prezentate în tabelul 9.2.

În raportul prezentat, dr. Wiley scrie : „Reorganizarea propusă pentru sistemul de rezervare va conduce la un serviciu mult mai nesatisfăcător pentru clienţi. Pentru a egala actualul standard de servire, care este suficient de bun, vor fi necesari cel puţin doi salariaţi în fiecare zonă, în timpul orelor de aglomeraţie. Aceasta va însemna o creştere de 100% a personalului. Odată cu creşterea traficului, sistemul propus devine şi mai nesatisfăcător decît cel actual, căci

creşterea numărului de angajaţi tinde să se facă cu multipli de 5 (numărul zonelor). Sistemul actual, cu toate telefoanele concentrate şi capabile să preia orice apel, admite o dezvoltare mai graduată odată cu creşterea traficului. Trabuie renunţat la ideea separării pe zone geografice şi trebuie reţinut din sistemul actual modul de

Tabelul 9.2

Comparaţie între sistemul existent şi cel nou

	Sistemul existent	Sistemul propus
Numărul de angajaţi	5	11
Procentul de apeluri ce găsesc toate liniile ocupate	1,3	1,6
Timpul mediu de aşteptare (min)	1,1	1,2
Procentul de apeluri ce aşteaptă mai mult de 1 min	31	39

grupare a telefoanelor. Dar pentru a face un sistem de rezervare mai eficient, este de dorit să se revadă modul de obţinere a informaţiilor de către angajaţi şi de asemenea modul de păstrare a informaţiilor".

Raportul a fost prezentat. Argumentele sînt bine primite : cu toate că trebuie făcut ceva în organizarea sistemului de rezervare, centrul de rezervare este menţinut la fel cum a fost ; ei bine, să zicem că aproape cum a fost.

Dr. Wiley, cu sentimentul că a făcut o treabă bună, pleacă de la şedinţă, de la Companie şi din această carte.

9.3. INTERLUDIU

Timpul trece şi au loc schimbări în sistemul de transport. Anterior departamentul era împărţit din punct de vedere funcţional în sectoarele : contabilitate, preţuri, documentaţie şi vamă-expediţie. Anumiţi angajaţi aveau sarcina de a ţine contabilitatea încasărilor, precum şi de a înregistra telefoanele prin care se cereau informaţii sau se anula vreo comandă. Această împărţire nu mai este acum valabilă. Departamentul de transport este împărţit acum în secţia de rezervări elefonice şi alte sectoare care se ocupă de preţuri, documentaţie,

vamă-expediție și urmărirea transporturilor cu o anumită destinație. Pe lîngă acestea, telefoanele prin care se solicită informații sînt preluate de către sectorul corespunzător zonei geografice respective.

Sistemul telefonic a fost de asemenea îmbunătățit. Anterior, orice angajat putea să răspundă la orice apel telefonic ce solicita informații de orice fel (adică era un sistem de informare complet). Acum cele cinci perechi de circuite telefonice interne de la centrala aeroportului sînt repartizate la cele cinci sectoare. Clienții permanenți au fost sfătuiți să ceară o anumită linie atunci cînd își rezervă locuri pentru o anumită destinație *.

De exemplu, dacă se solicită o rezervare pentru Geneva și dacă una sau ambele linii care deservesc sectorul 3 nu sînt ocupate, se conectează direct pe legăturile 31 sau 32. Dacă ambele sînt ocupate, apelul telefonic este conectat la orice altă linie liberă.

Acest sistem s-a dovedit nesatisfăcător pentru departament, deoarece clienții s-au plîns că secția de rezervări telefonice funcționează defectuos. Atunci s-a încercat să se amelioreze situația, reintroducîndu-se angajați în fiecare sector care să se ocupe de rezervările telefonice. Directorul îl însărcinează pe dl. Staple să întreprindă o nouă anchetă.

9.4. FAZA A DOUA

Din fericire Staple este un om cu vederi largi care verifică totul în practică. Cele zece linii telefonice din secția de rezervări vor fi și pe viitor înpărțite în cinci grupuri a cîte 2 linii, dar fiecare grup va recepționa numai acele apeluri telefonice care sînt strict legate de activitatea sectorului respectiv. Telefoniștii de la centrală vor reține acele apeluri care nu pot fi conectate la nici una din cele două linii corespunzătoare. În timpul orelor de vîrf, o colaboratoare va controla la întîmplare diferite linii pentru a vedea cum decurge activitatea. Telefoniștii de la centrală (care, deși comentează, sînt mulțumiți că participă la un proiect special) au fost de acord să numere apelurile telefonice dintr-o oră pentru fiecare sector. Dl. Staple a cronometrat și el la întîmplare aceste apeluri.

Experiența a durat două săptămîni, iar imensul volum de date strînse a fost analizat. Iată cîteva dintre rezultatele obținute.

* Centrala automată reprezintă încă un proiect viitor.

Din tabelul 9.3 rezultă că timpul mediu de servire al fiecărui apel telefonic este mai scăzut față de cel obținut cu 12 luni înainte, dar diferența nu este semnificativă, la nivelul de 5%.

Tabelul 9.3

Comparațle între timpul de funcționare și cel de așteptare

	Studiul I	Studiul II
Timpul mediu de servire al unui apel telefonic	3,7	3,7
Timpul mediu de așteptare al unui apel telefonic	1,35	1,58

Însă nu s-a mai răspuns la apelurile scurte care solicitau informații. Dl. Staple a tras concluzia în mod just că angajații, fiind specializați pentru o zonă geografică relativ restrînsă, pot să prelucreze mai repede apelurile telefonice care solicită rezervări. Timpul mediu de așteptare în orele de vîrf este mai mare decît era anterior, fapt datorat :

a) unui număr sporit de apeluri în timpul orelor de vîrf ;

b) unei discrepanțe mai accentuate între procentele apelurilor corespunzătoare diverselor sectoare ;

c) introducerii grupurilor cu acces limitat.

Este adevărat că numărul mediu al apelurilor telefonice care sosesc la centrală pe oră în timpul perioadelor de vîrf a crescut, atingînd circa 75 apeluri pe oră. Nu există nici un motiv să presupunem că repartiția lor nu este Poisson.

Procentele apelurilor pentru diverse sectoare înregistrate în timpul experienței au fost următoarele :

1) Insulele britanice 17% (17,0)
2) Europa de nord și est 22% (21,6)
3) Europa de sud 29% (28,6)
4) Orientul Apropiat 12% (12,2)
5) Africa 20% (20,6)

Numerele din paranteze reprezintă rezultatele analizei dr. Wiley. Aparent nu s-a produs nici o schimbare și putem să tragem concluzia,

la fel ca dl. Staple, că principala cauză a deteriorării serviciului o reprezintă faptul că nu există un acces complet.

Adepții teoriei așteptării sînt satisfăcuți, dar dl. Staple este impresionat de reducerea timpului de servire datorită specializării. Mai mult, el a mers mai departe decît dr. Wiley (care observase că angajații trebuiau să-și irosească o parte din timp cu căutarea informațiilor adecvate și sugerase să se facă ceva în această direcție) și a efectuat măsurări care au arătat că pentru acele convorbiri telefonice care nu-i obligă pe angajați să își părăsească locul, timpul mediu de servire pentru un apel este de 2,3 minute față de 3,5 minute în general. Dl. Staple consideră că dacă rezervările pentru diferite zboruri s-ar face telefonic de către angajații împărțiți pe sectoare, atunci informațiile privind aceste sectoare pot fi păstrate în stative special amplasate între cele două posturi telefonice care pot recepționa apelul telefonic. Ambii angajați pot ajunge cu ușurință la ele fără să-și părăsească locul. El face o nouă paralelă, avînd în vedere trei sisteme posibile :

a) aceiași angajați se ocupă de rezervări și de informații privind toate zonele geografice, avînd acces la toate liniile telefonice ;

b) un grup de angajați se ocupă de rezervări pentru toate zonele geografice, informațiile însă sînt date de angajați speciali în fiecare sector în parte ;

c) rezervările și informațiile pentru fiecare sector sînt date separat de către angajații sectorului corespunzător.

Timpul mediu de servire care este luat în considerare în calcule este dat în tabelul 9.4.

Tabelul 9.4

Timpii de servire

Sistemul	(a)	(b)	(c)
Apeluri telefonice pentru rezervări	3,5	3,5	2,3
Apeluri telefonice pentru informații	3,2	2,8	2,8

Aceste date se bazează pe observații directe în cazul sistemelor (a) și (b) și pe o ipoteză suplimentară în cazul sistemului (c). Această ipoteză constă în faptul că buna cunoaștere a informațiilor de către angajații care se ocupă de rezervări într-un anumit sector reduce durata medie a convorbirilor pentru rezervări la cea a apelurilor

telefonice pentru rezervări de care se ocupă un angajat în sistemul (a) atunci cînd acesta nu trebuie să-şi părăsească biroul (adică nu pierde timp să caute informaţiile cerute în altă parte).

Sistemul (a) are meritul că fiecare angajat poate răspunde la apelurile pentru orice zonă, dar presupune o risipă mai mare de timp din partea angajaţilor care trebuie să-şi părăsească locurile pentru a consulta informaţiile. Sistemul (c) impune funcţionarilor să-şi părăsească locurile mai puţin dar impune o diferenţiere mai mare a tipurilor de apeluri şi întîrzieri datorite transferului apelurilor dintr-un sector în altul. Acest fapt este important deoarece 21% dintre apelurile pentru rezervări includ tranzacţii multiple (doar 10% în 1957). Sistemul (b) este un compromis. Pentru fiecare sistem a fost calculat (tabelul 9.5) numărul personalului necesar în timpul unei ore de vîrf pentru a asigura servirea clienţilor (măsurat cu timpul mediu de aşteptare de circa 1 min/apel). Sistemul (b) este mai puţin recomandabil şi alegerea se va face între sistemul (a) şi sistemul (c).

Concluziile doctorului Wiley au fost răsturnate deoarece este evident că cea mai bună soluţie este ca angajaţii care se ocupă de rezervări telefonice să fie amplasaţi în sectoarele respective. De fapt dl. Staple a folosit acelaşi model ca şi dr. Wiley, modelul clasic

Tabelul 9.5

Necesar de personal în orele de vîrf

Sistemul	(a)	(b)	(c)
1. Angajaţi care lucrează permanent (care nu se ocupă decît de convorbirile telefonice)	7	5	—
2. Angajaţi care lucrează un timp limitat	—	Sectoarele 2, 3 : *2* Sectoarele 1, 4, 5 : *1* Total = 7	Sectoarele 2, 3 : *3* Sectoarele 1, 4, 5 : *2* Total = 12
3. Folosirea angajaţilor permanenţi	80%	70%	—
4. Folosirea angajaţilor pe timp limitat	—	40%	38%

M/M/c : (∞/FIFO). Deosebirea a apărut deoarece dl. Staple şi-a dat seama că sistemul poate fi ameliorat inventînd un mijloc de reducere a timpului de servire, lucru evident dar, de cele mai multe ori, neglijat!

9.5. COMENTARIU

Atit Wiley cît şi Staple au dreptate şi se înşeală. Wiley a preferat (în mod just) metoda servirii pe mai multe canale specializării înguste, iar Staple a insistat corect asupra scurtării timpului de servire. Cititorul va lua în consideraţie ambele concluzii. Nu este foarte greu să se inventeze un sistem de vizualizare, sau bazat pe computere, care să permită funcţionarilor de la rezervări, care se află într-o sală, să răspundă la toate întrebările şi să facă toate operaţiile de rezervare repede şi precis fără să-şi părăsească locurile. Această soluţie a fost adoptată pe scară largă acum, după mulţi ani, de către multe companii aeriene.

În limbajul capitolului 7, configuraţia optimă a timpului de aşteptare este influenţată de anumiţi factori — în acest caz, de timpul mediu de servire al apelurilor telefonice. Pentru o anumită valoare, soluţia lui Wiley a fost cea mai bună; pentru alta cea a lui Staple. Staple a făcut aproximări cam grosolane folosind tabelele lui Molina pentru $M/M/c:(\infty/\text{FIFO})$ pentru a deduce valorile aproximative din tabelul 9.5. Ar fi fost mai bine dacă ar fi fost folosite modelele $M/M/c:(20/\text{FIFO})$ şi $M/M/c:(5/\text{FIFO})$ ca mai bune ilustrări ale comportării viitoare a sistemului. Deşi rezultatele estimative ale oamenilor diferă, totuşi clasificarea celor trei sisteme posibile este aceeaşi.

Această concluzie a mai fost trasă în cap. 8.

10

METODE DE CONTROL AL CĂLĂTORILOR

Al doilea caz pe care-l vom examina este sistemul de verificare a pasagerilor la aeroport. Majoritatea cititorilor cunosc sistemul de verificare, iar pentru cei care nu sînt familiarizaţi vom da o descriere a acestei operaţii.

Atunci cînd un pasager soseşte la aeroport să ia avionul, mai întîi trebuie să se prezinte la un ghişeu de verificare unde angajatul liniei aeriene, numit şi agent de pasageri, cîntăreşte bagajele călătorului, înregistrează rezultatul cîntăririi pe talonul de zbor al călătorului (acea parte a biletului care indică despre care zbor este vorba), etichetează bagajele călătorului, le trimite la camera de bagaje, îi dă călătorului contramarca bagajelor şi, în unele cazuri, îi ia cuponul de zbor şi îi dă în schimb un permis de trecere. După ce au fost îndeplinite toate aceste operaţii, călătorul este admis pentru zbor.

Cu excepţia aeroporturilor mici, în anumite perioade ale zilei verificarea călătorilor pentru mai multe zboruri se face în acelaşi timp şi compania aeriană trebuie să asigure mai multe ghişee care să servească pasagerii. Există numeroase moduri de a organiza aceste ghişee pentru a asigura o servire eficientă, însă majoritatea celor folosite în mod curent sînt amestecuri judicioase (mai mult sau mai puţin) a două tipuri extreme. Primul tip poate fi denumit sistem cu restricţie completă sau sistem de zbor. În această formă de organizare fiecărui zbor îi este distribuit un grup de ghişee (adesea unul singur). Toţi pasagerii pentru un anumit zbor trebuie să fie verificaţi la ghişeele special desemnate, care nu se ocupă decît de pasagerii acestui zbor. Cel de-al doilea tip este cunoscut fie ca sistem liber, sau comun, sau ca sistem cu acces liber. În acest sistem fiecare

pasager care soseşte poate să fie verificat la orice ghişeu al companiei aeriene.

Primul sistem are dezavantajul că pasagerii aşteaptă de regulă la un ghişeu în timp ce altul, repartizat pentru un alt zbor, este liber; se poate admite ca acest fenomen nu apare în cazul unui sistem cu acces liber. Lipsa principală a sistemului cu acces liber este că funcţionarul de la ghişeu nu ştie pentru ce zboruri stau la coadă pasagerii în faţa ghişeului său. Călătorul care vine tîrziu pentru un anume zbor poate sta la coadă după alţii care sînt verificaţi pentru zboruri ulterioare şi astfel poate întîrzia pînă la a pierde avionul; funcţionarul nu poate să anticipeze asemenea situaţii.

În anii 1950, în Anglia, liniile aeriene foloseau exclusiv metoda tradiţională de verificare şi nu aveau experienţă în ceea ce priveşte caracteristicile operaţionale ale sistemelor obişnuite. În 1955 s-a prevăzut că creşterea rapidă a traficului aerian şi introducerea iminentă a avioanelor mai mari face ca sistemul de verificare să fie neeconomic şi incomod, cel puţin pentru aeroporturile mari şi aglomerate.

10.1. PRIMA FAZĂ

În 1956 secţia de cercetări operaţionale a companiei BEA a întreprins un studiu asupra sistemului de verificare a călătorilor la aeroportul din Londra. Acest studiu avea un scop dublu: să stabilească în mod obiectiv cum funcţionează sistemul de verificare existent şi să-i stabilească perspectivele viitoare. Sistemul de verificare existent era un sistem de tip mixt. Toţi pasagerii care soseau să fie verificaţi cu mai mult de o oră înainte de plecare erau serviţi de către unul dintre cele două ghişee speciale de înregistrare. Pasagerii care soseau mai tîrziu erau verificaţi la celelalte ghişee. Înregistrările amănunţite ale momentelor sosirilor pasagerilor au arătat că aproximativ 30% din totalul călătorilor au fost verificaţi la ghişeele speciale de înregistrare. Figura 10.1 prezintă graficul sosirii călătorilor exprimat prin proporţia $F(z, t)$ a călătorilor sosiţi în diferite momente pînă la ora închiderii (adică pînă la cel mai tîrziu moment posibil înainte de plecare) pentru verificare. După închidere nu se mai verifică nici un călător (zborul a fost „închis").

Din fig. 10.1 reiese că 70% din călători au fost verificaţi la alte ghişee. De fapt călătorul care soseşte mai tîrziu decît cu o oră înainte de ora zborului său, dar cu 25 min înainte de zbor, poate să fie verificat la un ghişeu special purtînd numărul zborului său. (Cititorul va observa că compania aeriană, în dorinţa de a-şi satisface clienţii,

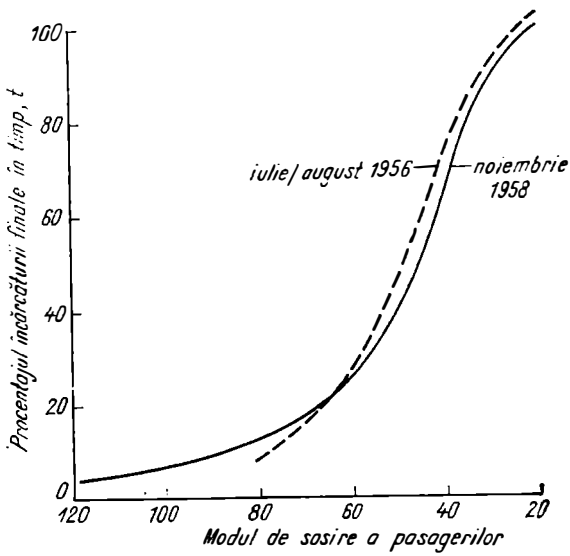


Fig. 10.1. Modul de sosire a pasagerilor.

verifică şi călătorii care sosesc mai tîrziu de ora închiderii ghişeului, adică cu 25 min înainte de zbor. Cu toate acestea, principiul rămîne în vigoare.) Figura 10.2 indică faptul că clienţii sosesc de obicei în grup organizat. Ca urmare, ca unitate de măsură a intrărilor s-a luat grupul de călători.

S-a examinat modul de intrare în sistemul de aşteptare la ghişeele speciale de înregistrare. Numerele grupurilor de călători care au sosit la interval de cinci minute pe parcursul celor două ore de vîrf ale celor cinci zile luate sub observaţie sînt prezentate în tabelul 10.1.

Din experienţă, este de aşteptat ca intrările să nu fie repartizate Poisson; în fond, au fost observaţi călători provenind de la un număr fix de surse (zboruri) care nu au toate aceleaşi efecte. Cititorul trebuie

să-şi reprezinte în mod clar repartiţia totală obţinută prin suprapunerea mai multor modele de tipul indicat în fig. 10.1, cu momentele de „închidere" a zborurilor decalate cîteodată cu 5 şi alteori chiar

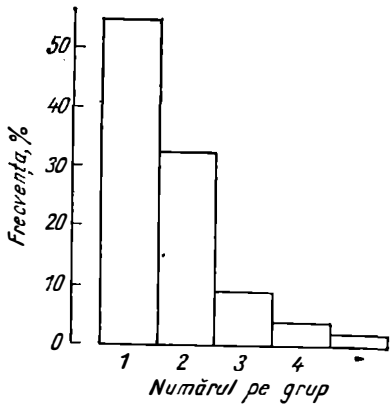


Fig. 10.2. Repartiţia numărului pasagerilor pe grup.

cu 10 minute unul de altul. Tabelul 10.1 indică faptul că totuşi ipoteza că intrările sînt repartizate Poisson nu este prea absurdă.

Evident, n-am avut un motiv pentru a face un test statistic de semnificaţie a acestor date, căci ştiam că sosirile nu pot fi în

Tabelul 10.1

Repartiţia numărului grupurilor de pasageri care au sosit la interval de 5 min la ghişeele speciale de înregistrare (ore de aglomeraţie)

Nr. grupurilor pe interval	Frecvenţa observată (%)	Repartiţia Poisson teoretică	Nr. grupurilor pe intervail	Frecvenţa observată (%)	Repartiţia Poisson teoretică
0	7	3,3	6	7	7,2
1	12	11,3	7	4	3,5
2	20	19,3	8	3	1,5
3	24	21,9	9	2	0,6
4	12	18,6	10	1	0,2
5	7	12,6	11	1	(0,05)
			Media	3,4	3,4

mod real repartizate Poisson. Altfel spus, repartiţia intrărilor, deşi nu este apriori Poisson, poate fi totuşi aproximată practic printr-o repartiţie Poisson.

A fost observată repartiţia timpului de servire la ghişeele speciale de înregistrare în timpul orelor de aglomeraţie (fig. 10.3). Nu poate

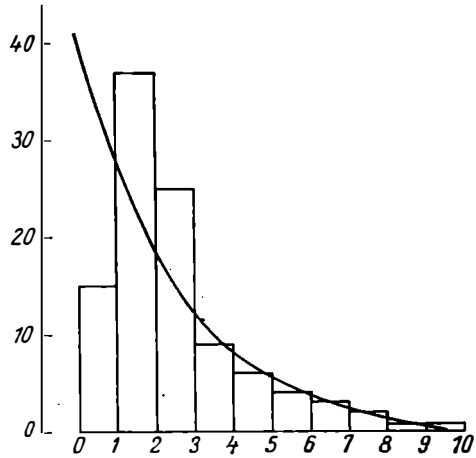


Fig. 10.3. Repartiţia timpului necesar prelucrării grupurilor de pasageri.

fi repartiţie exponenţială dar, pe de altă parte, repartiţia timpului de servire chiar şi în teoriile clasice privind traficul telefonic nu se asemăna prea mult cu repartiţia exponenţială. Grupul de cercetători a admis că repartiţia timpului de servire este exponenţială. Trebuie să se sublinieze faptul că repartiţia timpului de servire pentru ghişeele obişnuite de verificare a fost identică cu cea pentru ghişeele speciale de înregistrare. Aceasta nu este surprinzător, deoarece activitatea cerută pentru grupuri de pasageri este identică pentru ambele tipuri de ghişee. Figura 10.3 este compusă din suma datelor obţinute de la toate ghişeele. În acelaşi timp cu înregistrarea timpului de servire şi a orelor de sosire a călătorilor, cercetătorii au înregistrat timpul de aşteptare al călătorilor la cele două ghişee speciale de înregistrare. Aceste observaţii au fost verificate şi în alte perioade, nu neapărat de aglomeraţie. Apoi, timpii de aşteptare înregistraţi, în perioade de 1/2 oră, au fost reprezentaţi grafic în raport cu utilizarea

funcționarilor în aceleași perioade. Figura 10.4 prezintă aceste rezultate (cititorul își va aminti că $u = 100\rho$, v. cap. 4). Sînt sigur că și cititorii consideră aceste rezultate ca fiind încurajatoare. Observațiile empirice sînt în concordanță cu previziunile modelului teoretic $M/M/2 :(\infty/\text{FIFO})$ și cercetătorii au considerat, în mod just, că în analiza ulterioară toate ghișeele trebuie tratate ca și cum ar fi un sistem cu mai multe stații de servire în paralel, sistem ce poate fi studiat cu ajutorul modelelor de tipul $M/M/c : (\infty/\text{FIFO})$. Această extrapolare îndrăzneață a experienței s-a dovedit a fi greșită.

Sistemul de așteptare la ghișeele obișnuite la care au fost serviți călători în proporție de 71% este foarte neobișnuit. Cercetătorii nu au considerat rentabil să-l analizeze folosind modele matematice. Observațiile empirice și experiența dobîndită de ghișeele obișnuite au stabilit cîteva lucruri importante. Mai întîi, datorită ritmului variabil al sosirilor pasagerilor în timpul perioadei de verificare (de la 60 la 25 min înainte de momentul plecării), utilizarea unui singur funcționar la un asemenea ghișeu a fost foarte variabilă (tabelul 10.2).

Tabelul 10.2

Utilizarea funcționarilor la ghișeele de verificare

Perioada min înainte de plecare	$u(\%)$
60 — 50	53
50 — 40	53
40 — 31	76
30 — 25	15

În timpul perioadei de 40—31 minute înainte de plecarea cursei, de obicei este necesar ca un al doilea funcționar să vină să ajute; cei mai experimentați funcționari chemau automat în ajutor un asemenea auxiliar cu 45 min înainte de plecare pînă cu 25 min înainte de plecare (în timpul acestor ultime 5 min, 30—25 înainte de plecare, deși erau puține sosiri ale călătorilor, cele care se înregistrau erau

foarte critice; în special în timpul ultimelor sosiri de după ora de închidere, al doilea funcţionar nu era deloc în plus).

Cu ajutorul acestor stratageme, funcţionarii de la ghişeele obişnuite de verificare au reuşit să menţină timpul de aşteptare al călătorilor la un nivel tolerabil, comparabil cu cel obţinut la ghişeele speciale

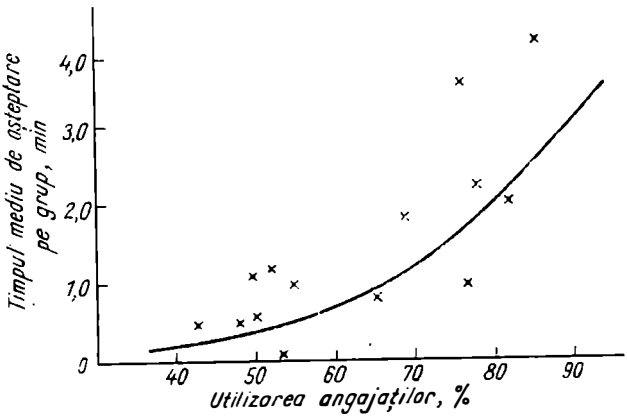


Fig. 10.4. Creşterea timpului de aşteptare a pasagerilor la ghişeele speciale de înregistrare odată cu creşterea ocupării angajaţilor.

* Observaţii efective se bazează pe medii calculate pe perioade de 1/2 oră – Curba dată de $M/M/2$; (∞/FIFO).

de înregistrare. Timpul mediu de aşteptare în orele de vîrf la ambele tipuri de ghişee varia între 1,2 şi 2,5 minute. S-a pus întrebarea: este această valoare satisfăcătoare?. Conducerea companiei a considerat că nu şi s-a recomandat ca standard de calitate ca timpul mediu de aşteptare în orele de aglomeraţie să fie de cel mult un minut.

O altă problemă care se pune: este posibil ca acest sistem să fie extins pentru a face faţă unui volum sporit de călători? Se părea că ghişeele speciale de înregistrare nu pun nici o problemă. Extinderea la trei sau patru va asigura acoperirea solicitărilor pentru următorii cinci ani*.

Se estima dublarea numărului călătorilor între anii 1956 şi 1961. Unul din principalii factori care determină această creştere îl constituie introducerea avioanelor de capacitate sporită. Desigur, se estima

* Compania BEA era într-o perioadă de dezvoltare rapidă.

că va creşte şi numărul zborurilor, dar nu cu mult, însă creşterea numărului de pasageri trebuia să aibă loc în special prin înlocuirea vechilor tipuri de avioane cu tipuri noi, de capacitate mult mai mare decît cele vechi. De exemplu, avioanele de tipul Vickers Vanguard au o capacitate dublă faţă de tipul mai vechi Viscount 701S. Acest lucru va influenţa considerabil situaţia la ghişeele obişnuite. Astfel, conform calculelor, cu 40—25 min înainte de plecarea avioanelor mari, vor fi necesari pînă la trei funcţionari pe zbor, iar cu 60—40 min înainte de zbor cîte doi, pentru a se evita aglomeraţia, lucru care a fost considerat inadmisibil.

Una dintre soluţiile propuse a fost de a scurta timpul de transfer de la ghişeele speciale, de la 60 min înainte de plecare la 40 min înainte de plecare, astfel încît ghişeele obişnuite să fie afectate doar cîte 15 min unui zbor. Examinîndu-se această soluţie, s-a constatat că se poate economisi timp dirijîndu-se cît mai mulţi călători spre ghişeele speciale de înregistrare. În acest scop s-au folosit ca modele $M/M/5$: (∞/FIFO) pentru ghişeele speciale şi $M/M/2$: (∞/FIFO) pentru perechile de ghişee obişnuite. Ultimul model este evident foarte grosolan.

Înainte ca acest sistem să poată fi pus în practică trebuiau să se rezolve cîteva probleme de funcţionare. Deoarece sistemul de verificare existent nu presupunea mai mult de doi angajaţi — sau ghişee — pentru fiecare zbor, nu a existat o problemă a controlului. Angajatul (angajaţii) putea(u) să verifice pasagerii urmărind lista, obţinîndu-se un control strict al numărului şi identităţii acestora. Este clar că în cazul noului sistem apare necesitatea confruntării biletului într-un punct central pentru toţi pasagerii unui zbor, deoarece un număr mai mare de angajaţi nu poate utiliza aceeaşi listă, iar identitatea moment cu moment a mai multor liste este greu de realizat. Transferul biletelor de zbor într-un punct central de control pe o bandă transportoare sau prin poştă pneumatică generează un alt şir de aşteptare la camera de control, unde biletele ar mai trebui să fie sortate în funcţie de destinaţia zborului.

Centrul de control ar trebui să îndeplinească deci următoarele funcţii : să sorteze biletele după indicativul zborului pe măsură ce acestea sosesc pe bandă de la ghişeele obişnuite; să ia fiecare set de bilete la închiderea zborului respectiv, să adune greutatea bagajelor şi să stabilească numărul pasagerilor, să comunice aceste date secţiei de control al încărcăturii, să verifice biletele după lista pasagerilor şi să observe neconcordanţele (pasageri care nu s-au prezentat,

alţi pasageri suplimentari etc.); cînd este necesar, să expedieze datele care prezintă interes la aeroportul de destinaţie. Unele dintre aceste operaţii erau efectuate de secţia de control, iar în sistemele de verificare existente nu erau necesară sortarea (biletele erau deja clasificate), adunarea greutăţii bagajelor era făcută de către funcţionarii de la ghişee, iar verificarea biletelor pe baza listei pasagerilor la fel.

Trebuia să se evalueze mîna de lucru necesară pentru îndeplinirea tuturor acestor operaţii în noul centru de control. Astfel, a fost stabilit timpul mediu de servire care va rezulta pentru ghişeele obişnuite dacă o parte din aceste operaţii vor fi transferate spre centrul de control. S-a prevăzut că va fi în medie de 1,6 minute pentru un grup de călători şi s-a convenit ca repartiţia sa să fie considerată exponenţială. De asemenea s-a calculat numărul de angajaţi necesari în centrul de control, sintetizîndu-se şi calculîndu-se volumul total de muncă, aplicîndu-se un factor de utilizare. S-a apreciat că în centrul de control este nevoie de n persoane, unde n a fost calculat după cum urmează:

$n = 2 + 0{,}2 \times$ (numărul funcţionarilor de la ghişeele obişnuite) (10.1)

Pînă acum studiul s-a bazat pe fapte. Următoarea problemă pusă ţinea mai mult de imaginaţie: ce fel de disciplină a aşteptării se va obţine în cazul unui sistem de verificare obişnuit? Răspunsul era greu de dat. După cum vă amintiţi, cele două ghişee speciale de verificare au funcţionat ca şi cum s-ar fi aplicat regula FIFO. Părea valabil ca pasagerii să se răspîndească la ghişee, deplasîndu-se liber de la unul la altul, pentru a-şi reduce timpul de aşteptare. Deşi nu se respectă strict disciplina clasică de aşteptare FIFO: clienţii sînt serviţi în ordinea sosirii lor, totuşi s-a apreciat că se întîmplă aproximativ acest lucru. Standardul de servire luat drept bază de calcul era un timp mediu de aşteptare care nu este influenţat de disciplina de aşteptare în modelele particulare cu mai multe staţii de servire care păreau că aproximează sistemul real de verificare; cercetătorii au fost de acord să admită regula FIFO ca regulă a disciplinei de aşteptare.

Astfel, ca model pentru sistemul de verificare la ghişeele obişnuite a fost stabilit $M/M/c$: (∞/FIFO) cu $\mu = 0{,}625$ şi λ (în orele de aglomeraţie) iniţial 3,1 pe minut, momentul de transfer fiind cu 40 min înainte de plecare. Graficele lui Molina arată că pentru a ajunge la un timp mediu de aşteptare de 1,0 min în aceste condiţii, numărul de angajaţi trebuie să fie $c = 6$. În mod similar, c poate fi calculat pentru alte valori ale lui λ care corespund altor ore ale zilei.

Numărul ghișeelor de zbor, numite „ghișee finale de verificare", a fost stabilit pentru fiecare oră, determinîndu-se numărul de zboruri maxim care pleacă în timpul fiecărui interval de 15 min (fiecare ghișeu final trebuie să fie deschis timp de 15 min) și repartizîndu-se un angajat dacă avionul este de tipul DC—3, 1,5 pentru un avion de tipul Elisabethan sau Viscount, 2 pentru avion de tipul DC—6B. Astfel a fost calculat numărul personalului de control.

Necesarul de angajați și ghișee astfel calculat a fost ceva mai mic decît cel necesar pentru sistemul vechi de verificare a zborurilor; și astfel personalul a început să fie sigur că se va adopta noul sistem de verificare cît mai curînd posibil.

În acest punct al studiului s-a întîmplat ca cinci membri ai personalului de verificare să se îmbolnăvească cam în același timp într-o perioadă foarte aglomerată a anului. Supraveghetorii au fost de acord că este imposibil să se prelucreze fluxul de călători cu vechiul sistem de verificare, fără personal suficient, și au hotărît să încerce noul sistem de verificare: un sistem fără ghișee finale. Principalele probleme ce apăreau au fost următoarele: timpul era prea scurt pentru instruirea personalului rămas și lipsea un sistem de transport al taloanelor într-un punct central de control și sortare; ultima problemă a fost rezolvată ad-hoc, folosindu-se drept curieri angajații tineri care duceau taloanele la centrul de control format în grabă.

Noul sistem de verificare și prototip au intrat în funcțiune cu două zile înainte de un week-end aglomerat specific englezesc. La început totul a decurs conform planului, dar pe măsură ce a crescut traficul, au apărut deficiențe în serie. Timpul de așteptare al clienților era cu mult mai mare decît cel anticipat. Principala problemă privea influența sistemului asupra punctualității plecării avioanelor. Mulți pasageri care soseau cu 5—6 min înainte de închiderea zborului nu aveau timp să ajungă la ghișee în timp util. Pe bună dreptate aceștia reclamau, blocînd ghișeele și întîrziindu-i și pe pasagerii din spatele lor, astfel ajungîndu-se la întîrzierea plecării mult mai multor zboruri decît în mod obișnuit. Apariția acestor blocări a fost observată de aproximativ 20 de ori, din care una a durat chiar 18 min. În a patra zi a experienței, supraveghetorii au inițiat anunțarea pasagerilor cu cinci min înainte de momentul de închidere a fiecărui zbor, astfel încît cei care nu au fost verificați încă să treacă imediat în față, cerînd prioritate de servire. Această măsură nu a avut rezultate semnificative pentru că englezii, care sînt foarte disciplinați, nu au vrut să treacă în fața altora, și în cazul în care

au făcut-o s-a produs dezordine : călcîiul lui Ahile s-a dovedit a fi problema bagajelor.

Importanţa acestui experiment rezidă în faptul că a scos la iveală problema călătorilor care sosesc tîrziu şi a celor cu probleme speciale. De asemenea s-a stabilit că momentul la care trebuie să se fi încheiat verificarea ultimului pasager (momentul de completare a zborului) nu trebuie să depăşească cu cinci min momentul de închidere, altfel în 25% din cazuri plecările avioanelor vor fi întîrziate. Cheltuielile şi neajunsurile provocate de aceste plecări întîrziate companiei aeriene sînt aşa de mari încît conducerea acestor companii nu acceptă sisteme de verificare care pot să le favorizeze. În acelaşi timp, conducerea (care a fost informată despre gradul de aglomeraţie constatat în orele de vîrf) a fost de acord să îmbunătăţească standardul de servire la ghişeele obişnuite ; de acum înainte, timpul mediu de aşteptare al călătorilor va trebui să fie cel mult 30 s în orele de aglomeraţie.

În acest moment era necesar să se redefinească problema sistemului de verificare optim. Acest sistem va fi cel care va necesita numărul minim de salariaţi astfel încît : 1) să se asigure un timp mediu de aşteptare de 0,5 min în orele de vîrf ; 2) să se asigure ca cel mult pentru un zbor din 20 momentul de completare să fie cu 5 min mai tîrziu decît momentul de închidere. În aceste condiţii stabilirea unui sistem de verificare optim este acelaşi lucru cu stabilirea timpului de transfer optim pentru un sistem compus. Aceasta a reprezentat baza pentru o nouă serie de studii.

10.2. FAZA A DOUA

Cercetătorii care au întreprins această a doua serie de studii aveau experienţă mai multă şi erau mai multilaterali. Ei au întreprins o examinare metodică a celor patru componente ale unui sistem de verificare compus :

1. Verificarea pasagerilor la ghişeele obişnuite. Aceasta cuprinde : cîntărirea şi etichetarea bagajelor, înscrierea greutăţii bagajelor pe taloanele de zbor, întocmirea unor chitanţe pentru eventuale plăţi suplimentare, emiterea de permise de trecere şi transmiterea taloanelor de zbor spre camera de control.

2. Verificarea călătorilor la ghişeele finale cuprinde toate operaţiile îndeplinite la un ghişeu obişnuit, plus totalizarea numărului de călători, plus a greutăţii bagajelor, împreună cu alte cîteva operaţii minore.

3. Sortarea taloanelor care sosesc pe bandă de la ghişeele obişnuite, în funcţie de indicativul zborului.

4. Controlul, adunarea şi tipărirea în mod automat a numărului călătorilor şi greutăţii bagajelor acestora, confruntarea taloanelor cu listele de călători etc.

Cercetătorii au hotărît să studieze acest proces complex în anumite condiţii specifice de trafic şi anume:

1. În timpul unei ore de aglomeraţie se va programa ca plecarea zborurilor să aibă loc din cinci în cinci minute.

2. Pentru fiecare zbor în parte, 20 de grupuri de pasageri vor sosi înainte de momentul de închidere care este cu 25 minute înainte de momentul plecării.

Trebuie remarcat faptul că numărul de pasageri dintr-un grup variază, după cum se indică şi în fig. 10.2, şi că verificarea pentru zbor nu se face numai la aeroport, deoarece unii pasageri preferă să fie verificaţi la sediul din oraş.

10.2.1. VERIFICAREA LA GHIŞEELE COMUNE

Modelul $M/M/c$: (∞/FIFO) al sistemului de aşteptare la ghişeele obişnuite nu a fost infirmat în practică. Totuşi, în timpul care s-a scurs de la studiile iniţiale, la un alt aeroport englez a intrat în funcţiune un sistem compus. Din nefericire, proiectul conceptual şi fizic al acestui sistem a fost atît de particular, încît experienţa dobîndită aici nu a putut fi extinsă, cu o singură excepţie, anume repartiţia timpului de servire. Activităţile executate la un ghişeu obişnuit la aeroport erau identice cu cele executate la aeroportul din Londra. Ca urmare repartiţia prezentată în fig. 10.5 a înlocuit repartiţia teoretică folosită anterior.

Dificultatea pe care o pune această repartiţie constă în faptul că se ştia prea puţin despre modelele de aşteptare cu mai multe staţii de servire avînd repartiţii ale timpului de servire de tip Pearson III. Cercetătorii, cel puţin în Anglia, sînt tentaţi să facă următorul raţionament. În cazul unui sistem cu o singură staţie de servire ştim că timpul mediu de aşteptare produs de o repartiţie a timpului de servire

de tip exponențial negativă este mai mare decît cel produs prin repartiții de dispersii mai mici, toate celelalte condiții fiind aceleași. Formula lui Pollaczek asigură aceasta. Este logic să se presupună că

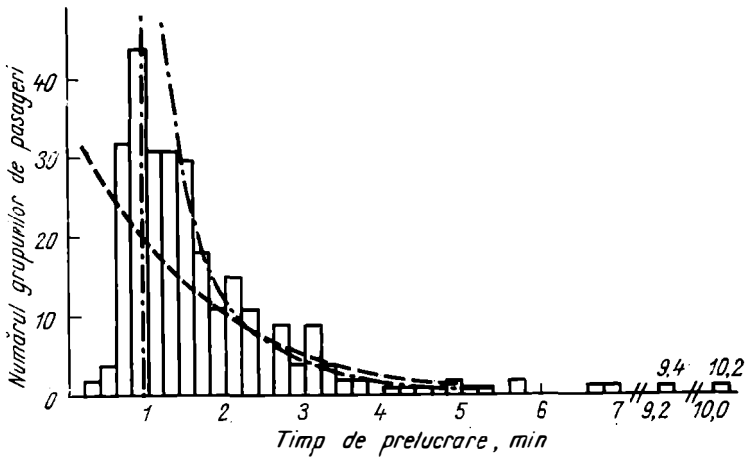


Fig. 10.5. Repartiția timpilor de verificare la ghișeele obișnuite. Dreptunghiurile reprezintă repartiția observată.
– . – . – . – Curba de tip Pearson III.
– – – – Curba exponențială negativă.

lucrurile se întîmplă la fel și în cazul mai multor stații de servire. Dacă calculăm totul ca și cum timpul de servire ar fi exponențial, acționăm cu prudență și conservatorism.

Un membru al echipei de cercetători englezi a construit o simulare modestă și a făcut-o să funcționeze manual (nu avea la dispoziție un calculator) pentru a stabili timpul mediu de așteptare care corespunde timpului de servire observat. El a comparat aceste rezultate cu cele obținute din modelul $M/M/c$:(∞/FIFO) cu ajutorul tabelelor lui Molina. Comparația s-a dovedit a fi atît de interesantă, încît rezultatele ulterioare au fost calculate, folosind diferitele repartiții Pearson III ale timpului de servire care au fost observate. Concluzia a fost că rezultatele foarte bune se pot obține repede calculîndu-se timpul mediu de așteptare pentru c stații de servire, fără a recurge la simulare, după formula

$$W_q^{(c)} = \frac{1}{2} M^{(c)} (1 + D^2), \qquad (10.2)$$

unde $M^{(c)}$ este timpul mediu de așteptare în cazul c stații de servire dacă repartiția timpului de servire ar fi exponențială negativă cu aceeași medie, și D este dispersia repartiției Pearson de tipul III. Această formulă a fost inițial obținută de Pollaczek pentru cazul unei singure stații de servire ($c = 1$). Concluzia la care s-a ajuns are o mare valoare deoarece este ușor de calculat $M^{(c)}$ din formula lui Erlang

$$M^{(c)} = P(0) \cdot \frac{(c\rho)^c}{c \cdot c!(1 - \rho)^2 \mu}, \tag{10.3}$$

unde $1/\mu$ este timpul mediu de servire iar

$$\frac{1}{P(0)} = \sum_{r=0}^{c-1} \frac{(c\rho)^r}{r!} + \frac{(c\rho)^c}{c!} \frac{1}{1 - \rho}, \tag{10.4}$$

după cum cititorul își amintește din cap. 3.

David Owen a sugerat o aproximare a timpului mediu de așteptare pentru un model din familia $M/E_k/c$:(∞/FIFO). În tabelul 10.3 sînt comparate rezultatele obținute, folosindu-se această formulă de aproximație cu cele obținute prin simularea Monte Carlo. Cele două serii de rezultate sînt în concordanță. Este clar că pentru repartițiile Pearson III ale timpului de servire care se întîlnesc în practică în sistemele de verificare (pentru care de obicei $0{,}70 \leqslant D \leqslant 1{,}0$) decizia de a se asigura c sau $c + 1$ stații de servire pe baza formulei lui Erlang corectată cu $\frac{1}{2}(1 + D^2)$ este suficient de corectă.

Această metodă de aproximație a rezolvat principala dificultate a calculelor și s-a putut calcula numărul ghișeelor, adică al stațiilor de servire necesare unui sistem compus de verificare pentru a prelucra fluxuri diferite de pasageri, asigurînd o servire standard echivalentă cu un timp mediu de așteptare de pasager de 0,5 min în orele de vîrf.

Cercetătorii au formalizat apoi estimația ritmului mediu de sosire a grupurilor de călători la ghișeele obișnuite. Presupunînd că sînt programate să plece în fiecare oră k zboruri la intervale de $60/k$ min, ritmul mediu de sosire trebuie să fie

$$\lambda = \frac{F(z_0, t)}{60} \cdot \sum_{i=1}^{k} n_i, \tag{10.5}$$

unde z_0 este momentul de transfer şi n_i numărul grupurilor de pasageri cu al i-lea zbor. $F(z, t)$ este proporţia cumulată a sosirilor cu z minute înainte de închiderea zborului, care are loc cu t minute înainte

Tabelul 10.3

Comparaţie între diferite evaluări ale timpului mediu de aşteptare

Nr. de staţii de servire	Cazul A		Cazul B		Cazul C	
	$b=2,50$ Formula de aproximare	$D=0,75$ Simulare	$b=1,64$ Formula de aproximare	$D=0,85$ Simulare	$b=1,59$ Formula de aproximare	$D=0,96$ Simulare
7	—	—	2,59	2,25	1,70	1,64
8	—	—	0,48	(0,27)	0,40	(0,51)
9	—	—	0,17	0,15	0,17	0,12
10	21,80	(8,4)	0,07	0,07	0,06	0,05
11	1,09	0,92	0,03	0,04	0,03	0,03
12	0,58	0,45	—	—	—	—
13	0,16	0,16	—	—	—	—
14	0,07	0,07	—	—	—	—

de închiderea zborului, care are loc cu t minute înainte de plecarea avionului.

Pentru valorile particulare date parametrilor modelului examinat, numărul ghişeelor de verificare necesare, corespunzătoare diferitelor momente de transfer, a fost găsit ca în tabelul 10.4.

Tabelul 10.4

Numărul angajaţilor necesari pentru ghişeele de verificare primite

Timp de transfer înainte de închidere, în min	5	10	15	20	25	30
c	8	8	8	7	6	5

La ghişeele de verificare obişnuite condiţiile legate de momentul completării unui zbor nu sînt importante decît dacă $z_0 - t = 0$ (adică dacă nu există ghişee finale) sau dacă probabilitatea ca un pasager să aştepte mai mult de $z_0 - t$ este considerabilă. O altă restricţie de care trebuie să se ţină seama : riscul ca un pasager care a sosit înainte de închiderea zborului să ajungă prea tîrziu la ghişeu.

Reformulînd aceste două observaţii prin luarea în consideraţie a unor limitări ale riscului, ele sună astfel :

1. Riscul ca un pasager care a sosit înainte de momentul de transfer să nu poată fi verificat înainte de închidere trebuie să fie atît de mic, încît să apară în cel mult 1 din 20 de zboruri.

2. Situaţia unei verificări a unui pasager începută înainte de închidere şi care nu s-a terminat după 5 min de la închidere trebuie să apară în cel mult 1 din 20 de zboruri.

Examinînd prima restricţie, s-a constatat că calculul riscului situaţiilor de acest tip implică cunoaşterea repartiţiei timpului de aşteptare. Pe cînd utilizarea modelului $M/M/c$:(∞/FIFO) apare justificată pentru calculul timpului mediu de aşteptare mai ales după corecţiile realizate folosindu-se ipoteza lui Owen, cercetătorii nu au avut tot atîta încredere că se vor obţine estimaţii corecte ale probabilităţilor asociate părţii superioare a repartiţiei timpului de aşteptare. Această dificultate a fost greu de rezolvat, deoarece se părea că nu era nici un mod de a se deduce repartiţia timpului de aşteptare pentru un model mai realist.

Simulările Monte Carlo făcute pentru un sistem obişnuit complet cu $z_0 - t = 0$ au generat totuşi repartiţii ale timpului de aşteptare, care sînt prezentate în fig. 10.6 pentru diverse numere de ghişee obişnuite. Este evident că într-un sistem compus în care momentul de transfer este z_0, probabilitatea $q(z_0, c ; v)$ a unui timp de aşteptare ce este mai mare decît v în condiţiile în care există c angajaţi, trebuie să satisfacă condiţia

$$q(z_0, c ; v) \leqslant q(t, c ; v),$$

unde t, ca de obicei, este momentul de închidere. Probabilitatea ca un pasager să sosească în intervalul $(u, u + \mathrm{d}u)$ este să spunem $p(u)\mathrm{d}u$; astfel încît probabilitatea necondiţionată ca el să sosească înainte de z_0 şi să nu ajungă în poziţia de a fi verificat pînă după momentul t este

$$Q\,(z_0, c) = \int_{z_0}^{\infty} p\,(u)\; q\,(z_0, c ; u - t)\,\mathrm{d}u \leqslant$$

$$\leqslant \int_{z_0}^{\infty} p(u)\; q(t, c ; u - t)\,\mathrm{d}u = Q'(z_0, c).$$

Probabilitatea ca cel puţin un grup de pasageri pe zbor să sosească înainte de momentul de transfer şi să nu ajungă în situaţia de a fi verificat decît după acest moment este mai mare decît

$$\{1 - [1 - Q'(z_0, c)]^{20}\}. \qquad (10.6)$$

Este uşor de arătat, folosindu-se repartiţiile timpului de aşteptare din fig. 10.6, că dacă această expresie ar fi mai mică decît 0,05, atunci

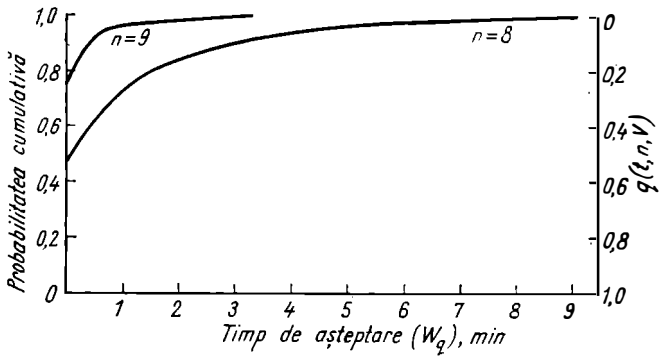


Fig. 10.6. Repartiţia cumulativă a timpilor de aşteptare pentru sisteme obişnuite complete de verificare ($z_0 = t$).

$z_0 - t$ trebuie să fie de cel puţin 7 min pentru 8 ghişee şi de cel puţin 2 min pentru 9 ghişee.

A doua condiţie a pus dificultăţi de calcul similare, care au fost rezolvate într-un mod analog. Pentru ca această condiţie să fie îndeplinită, s-a stabilit că momentul de transfer trebuie să fie cu cel puţin 2,6 min înainte de închidere. Rezultatele acestor calcule sînt :

1. Momentul de transfer nu va fi mai devreme de 3,6 min înainte de închidere.

2. Pentru un moment de transfer cuprins între 3,6 şi 7 min înainte de închidere, sînt necesare 9 ghişee.

3. Pentru momente de transfer situate mai devreme decît cu 7 min înainte de închidere numărul ghişeelor obişnuite este cel arătat în tabelul 10.4.

10.2.2. FUNCȚIILE CAMEREI DE CONTROL ȘI SORTARE

Cititorul își va aminti că în primul studiu, o regulă simplă bazată pe amestecarea și gruparea elementelor de timp de către un inginer a fost folosită pentru a se calcula raportul dintre personalul centrului de control și cel de la ghișeele de verificare. În timpul experimentului sortarea tichetelor și verificarea concordanței lor cu listele de călători s-a dovedit a fi o problemă serioasă. Acest fapt a sugerat că procesul sortării trebuie să fie examinat mai amănunțit. Mai mult, în primul studiu s-a presupus că operația de înregistrare a datelor va fi executată manual și că toate operațiile înrudite necesită un angajat pentru fiecare trei ghișee obișnuite. În perioada dintre primul studiu și experiment, unul dintre analiștii companiei a întreprins un studiu amănunțit al înregistrării datelor și al proceselor înrudite și ca rezultat a dezvoltat un nou procedeu de executare simultană a tuturor operațiilor aritmetice de trei ori mai repede și cu precizie sporită. Principala inovație care a făcut acest lucru posibil o reprezintă o modificare specială a unei mașini comerciale de înregistrat. Pentru a se aprecia avantajele acestei mașini, cercetătorii împreună cu analistul în cauză au făcut o serie de experiențe cu o machetă a camerei de control folosind un prototip al noii mașini de înregistrare.

Procedeul de sortare a tichetelor are o serie de caracteristici interesante și de aceea îl vom examina îndeaproape în următorul capitol al acestei cărți. Viteza cu care erau sortate tichetele a fost calculată după cum urmează. S-au folosit tipuri de stative pentru taloane, prevăzute să fie folosite în viitor. S-a făcut o experiență pentru a se stabili timpul cheltuit pentru a se sorta 160 taloane spre 16 destinații; factorii au fost : caracterul angajaților ce verifică pasagerii (trei tipuri), tipurile de stativ pentru taloane (patru) și tipurile de zile (trei). Cel mai bun sistem de sortare a avut un timp mediu de sortare de 2,1s pentru un talon, cu un coeficient de variație mai mic de 5%.

Intrările în poziția de sortare au fost considerate de tip Poisson și viteza medie a sosirilor tichetelor a fost $1{,}5/a$ pe minut, unde 1,5 este valoarea medie a taloanelor pe grup de călători, iar a este timpul mediu dintre sosirile grupurilor de pasageri la ghișeele obișnuite, definit pentru diferite momente de transfer ca fiind

$$\left[\frac{F(z_0, t)}{60} \sum_{i=1}^{k} n_i\right]^{-1}.$$

Era clar că chiar dacă toate operaţiile de verificare ar fi fost îndeplinite la ghişeele obişnuite, astfel încît solicitarea în poziţia de sortare ar fi maximă, traficul (exprimat în erlangi) ar fi de numai

$$\varepsilon = \frac{360 \times 2{,}1}{2600} = 0{,}21$$

şi întîrzierea în sortarea tichetelor ar fi mică chiar dacă nu s-ar ocupa de sortare decît un singur angajat.

Timpul necesar pentru înregistrarea a n pasageri şi a bagajelor acestora, înscrierea greutăţilor bagajelor şi transmiterea totalurilor la secţia de control a încărcăturii a fost calculat în mod similar şi avea o medie de $0{,}5n + 4{,}5$ min cu foarte mici abateri.

Intrările în procedeul de înregistrare au fost regulate şi la momentul de închidere pentru fiecare zbor trebuia să fie efectuată o totalizare. Regularitatea intrărilor, precum şi limita superioară a timpului de servire de 6 min pe zbor (corespunzînd verificării a 30 de pasageri) au arătat că este nevoie de cel mult doi angajaţi pentru această operaţie. Fiecare dintre ei se va ocupa de cîte un zbor şi vor prelucra zboruri alternative. Un singur funcţionar ar fi de ajuns dacă momentul de transfer ar fi întotdeauna cu mult înainte de momentul închiderii. Numărul mediu al taloanelor pe un zbor, pentru un moment de transfer z_0 a fost de $30F(z_0, t)$, să zicem $30F_0$. Aşadar, timpul mediu de totalizare a fost

$$(0{,}05)\,(30F_0) + 4{,}5 = 1{,}5F_0 + 4{,}5 \text{ min.}$$

Pentru ca această expresie să fie mai mică de 5 min este clar că $F_0 < 0{,}3$ şi astfel $z_0 - t$ nu trebuie să fie mai mic de 46 min, după cum se arată în fig. 10.1. Totuşi, deoarece numărul pasagerilor care trebuiau serviţi pe zbor avea o repartiţie binomială, timpul de totalizare poate să varieze şi momentul de transfer la 46 min înainte de închidere ar putea produce cozi excesiv de lungi în orele de aglomeraţie.

Volumul de interferenţă asociat unei probabilităţi de 0,05 ca timpul de totalizare să depăşească 5 min apare acceptabil pentru scopuri practice. Prin urmare, z_0 a trebuit să fie stabilit astfel încît

$$P\{1{,}5\,F_0 > 0{,}5\} \leqslant 0{,}05.$$

Aceasta este acelaşi lucru cu aflarea celei mai mari valori a lui p care satisface relaţia în care apare o funcţie beta incompletă :

$$I_p(10,\ 21) \leqslant 0{,}05.$$

Consultîndu-se tabelele funcţiei beta incomplete întocmite de Pearson, se constată că p nu trebuie să fie mai mare decît aproximativ 0,193. Aşadar, din fig. 10.1 rezultă că $(z_0 - t)$ trebuie să fie mai mare de 52 min. Astfel sînt necesari doi angajaţi pentru înregistrare şi control cu excepţia momentelor de transfer cu cel puţin 52 min înainte de închidere, în care caz unul singur este suficient. În plus, este necesar cel puţin un angajat pentru a face confruntarea bagajelor chiar dacă nu sînt deloc ghişee obişnuite.

În acest moment cercetătorii au putut calcula numărul total al angajaţilor neceari pentru secţia ghişeelor obişnuite de verificare ale unui sistem compus de verificare, corespunzător diferitelor momente de transfer. Rezultatele sînt date în tabelul 10.5.

Tabelul 10.5

Numărul angajaţilor necesari la secţia ghişeelor obişnuite a sistemului de verificare

$z_0 - t$ în min	0	5	10	15	20	40	60
Nr. de angajaţi	12	11	11	11	10	6	4

Sistemul de ghişee finale de verificare a zborurilor este diferit de tipurile studiate în general în teoria aşteptării. În primul rînd, intrările sînt neobişnuite. Examinarea înregistrărilor individuale colectate în timpul studiilor anterioare a arătat că numărul grupurilor de pasageri, r, care sînt verificate la un ghişeu final, peste cele 20 de grupuri pe zbor prevăzute iniţial, va avea o repartiţie binomială

$$b(r, 20;\ 1 - F_0) = \binom{20}{r} (1 - F_0)^r F_0^{20-r},$$

unde $F_0 = F(z_0, t)$. Totuşi, în orice caz numărul sosirilor posibile este strict limitat. Pe lîngă această particularitate, este clar că orice sistem de aşteptare la un ghişeu final de verificare începe ca şi cum nu ar fi şir de aşteptare în momentul z_0. Durata sa exprimată în numărul grupurilor de pasageri serviţi va fi foarte scurtă, astfel încît se poate cu greu presupune că se ajunge în starea de echilibru statistic. Cercetătorii au hotărît că singurul mod posibil de a trata acest sistem este de a-l simula.

Modelul sistemului de așteptare la ghișeele limitate de verificare a fost studiat și el prin simulare. S-a încercat să se obțină soluții de tranziție prin mijloace pur matematice, dar această sarcină s-a dovedit a fi peste posibilitățile echipei de cercetare. S-au efectuat 100 de serii de simulări în condiții diferie și fiecare combinație de condiții a fost reprodusă de cel puțin 10 ori. Simulările s-au făcut ținînd seamă de condițiile :

1) $z_0 - t < 60$ min și în plus ghișeul limitat se deschide la momentul z_0 și se închide la momentul t ;

2) numărul pasagerilor verificați la ghișeul limitat are repartiția binomială $b(r, 20, 1 - F_0)$;

3) numărul angajaților de la fiecare ghișeu este de 1 sau 2.

S-a presupus că $z_0 - t = 60$ corespunde unui sistem de verificare limitat complet și că în acest caz toți pasagerii care sosesc înainte de momentul z_0, care reprezintă aici momentul de deschidere a zborului și nu momentul de transfer, se împrăștie în holuri sau la bar și se întorc la întîmplare în timpul perioadei de verificare (z_0, t). Astfel

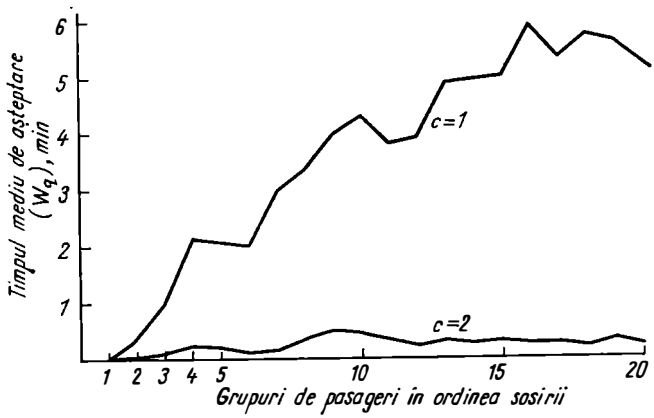


Fig. 10.7. Timpul mediu de așteptare al grupurilor de pasageri la ghișeele limitate ; rezultatele simulării în cazul $z_0 - t = 60$ min.

toate cele 20 grupuri de pasageri sosesc, din punctul de vedere al verificării, în timpul celor 60 min dinaintea închiderii fiecărui zbor. Rezultatele acestui caz (v. fig. 10.7) indică foarte clar că dacă va fi un singur angajat pe zbor, atunci timpul mediu de așteptare al pasa-

gerilor este complet inadmisibil pentru grupuri care ajung la ghișeu după al doilea și sugerează o nouă simulare pentru valori ale lui $z_0 - t$ mai mici cu aproximativ 5 min. Prin aceste mijloace cercetătorii au dedus că pentru un singur angajat (care în cazul nostru este stația de servire) transferul nu va putea avea loc mai tîrziu decît cu 6 min înainte de închidere.

Dacă sînt doi angajați pe zbor, orice moment este admisibil. Astfel, dat fiind timpul de așteptare standard a cărui medie va fi de aproximativ 0,5 min, un sistem limitat de verificare, complet, fără transfer, va fi acceptabil dacă sînt repartizați doi angajați pentru fiecare zbor timp de o oră.

Următoarea întrebare a fost în ce condiții un angajat va determina momente de completare acceptabile pentru $z_0 - t = 6$ min. Figura 10.8, dedusă din simulări Monte Carlo, prezintă repartiția

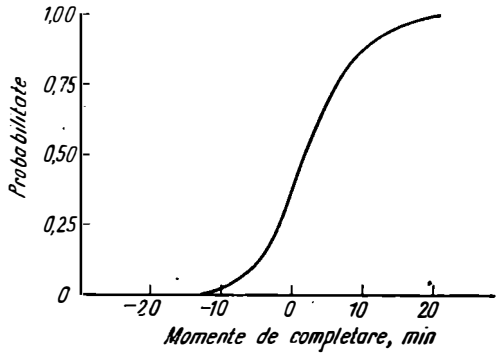


Fig. 10.8. Repartiția momentelor de completare la ghișeele finale, pentru transferuri ce au loc cu 6 min înainte de închiderea zborului ($z_0 - t = 60$); rezultate obținute prin simulare cu metoda Monte Carlo.

momentelor de completare pentru 100 de zboruri cînd transferul este planificat cu 6 min înainte de închidere. Chiar dacă selecția făcută este de volum mic, este clar că completarea are loc după 15 min de la închidere pentru mai mult decît proporția acceptabilă de 5 zboruri din 100.

În cele din urmă, cercetătorii au examinat problema dacă doi angajați pe zbor pot să aibă drept rezultat momente de completare

acceptabile şi, dacă este aşa, pentru ce valori ale momentelor de transfer. Figura 10.9 prezintă probabilitatea ca un moment de completare să depăşească 5 min, ca funcţie a momentului de transfer. Această figură sugerează că doi anagajaţi sînt suficienţi pentru momente de transfer nu mai tîrziu de 13 min înainte de închidere. De fapt, s-a hotărît să se accepte 15 min. Pe baza acestor rezultate s-a alcătuit tabelul 10.6, prezentînd numărul angajaţilor necesari pentru ghişeele limitate de verificare, într-un sistem compus, pentru diferiţi timpi de servire.

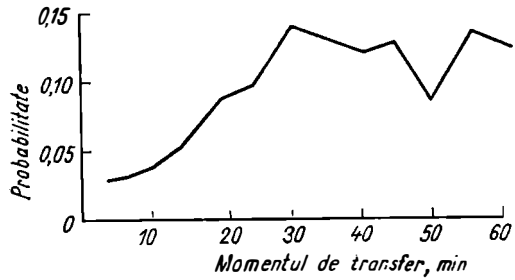


Fig. 10.9. Probabilitatea unui moment de completare întîrziat la ghişeele finale cu doi angajaţi; rezultate obţinute prin simulare cu metoda Monte Carlo.

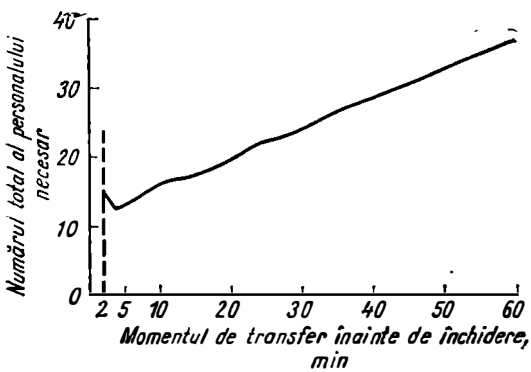


Fig. 10.10. Momentul optim de transfer.

Numărul angajaţilor ceruţi de sistemele de verificare compuse, pentru diferite momente de transfer, a fost obţinut însumîndu-se rezultatele pentru fiecare componentă. Rezultatele sînt prezentate în fig. 10.10.

Tabelul 10.6

Numărul angajaţilor la ghişeele finale

Timp de servire, min	0	5	10	15	20*	40*	60*
Nr. angajaţi	0	2	4	6	10	20	30

* Un angajat în plus la fiecare două zboruri pentru a satisface cerinţele momentelor de completare.

În practică este greu să se folosească un sistem compus de verificare cu momente de transfer care nu sînt multipli de 5 min. Astfel, figura 10.10 indică ca sistem optim acel sistem compus, în care transferul are loc cu 5 min înainte de închidere. În toamna lui 1958 noul sistem de verificare a fost inaugurat.

10.3. STUDIUL DE VALIDARE

Sistemul de verificare nu a satisfăcut pe deplin. Numărul de călători care s-au prezentat la ghişee a fost foarte apropiat de cel previzionat. Ghişeele au fost dotate cu personal în conformitate cu aceste previziuni, aşa cum se stabilise cu nouă luni în urmă. Dar cozi lungi tot mai erau şi, destul de ciudat, în special în perioadele relaxate ale zilei. Se părea că implementarea nu era corectă.

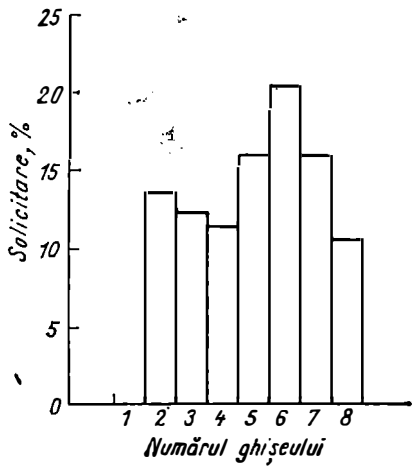


Fig. 10.11. Solicitarea diferenţiată a ghişeelor obişnuite de verificare.

Membrii echipei de cercetare, puţin modificate, au început să verifice schema sosirii călătorilor pentru zboruri. Diferenţa dintre aceasta şi cea care apărea cu doi ani în urmă a fost mică (v. fig. 10.1). Verificări statistice separate ale repartiţiei timpului dintre sosiri, cu

toate zborurile cumulate, au arătat că această schemă este exponențială negativă, la fel ca în momentul original.

Următoarea etapă a fost să se urmărească disciplina de așteptare. A fost măsurată solicitarea fiecărui ghișeu dintre cele opt, cu excepția primului care nu funcționa atunci. Figura 10.11 prezintă rezultatele : ghișeul 6 a fost aproape dublu solicitat față de nr. 8 și nr. 4. Călătorii nu se răspîndesc egal la toate ghișeele (în primul studiu s-a presupus că o fac). De ce? Explicația este dată de planul suprafeței de verificare prezentat schematic în fig. 10.12. Pasagerii merg la ghișeele cele mai apropiate de ușile prin care au intrat : ghișeul 2 a fost mai puțin solicitat datorită ușii duble prin care se ajunge la el pe calea cea mai scurtă, totuși apropierea de ușă l-a ajutat puțin.

Membrii echipei au observat că pasagerul care s-a așezat la un ghișeu în general nu se mai mută la altul. Au fost numeroase cazuri

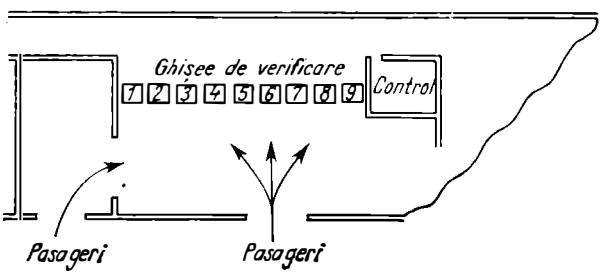


Fig. 10.12. Planul suprafeței de verificare în clădirea aeroportului, cu poziția ghișeelor și a ușilor de intrare. Ghișeele 1 și 9 erau folosite rar sau deloc.

cînd era coadă la ghișeul 7 și nici o persoană la ghișeele 4 și 3, pasagerii rămînînd la locurile lor. Acest lucru nu se întîmplă însă întotdeauna. Dacă vor să se deplaseze, călătorii o fac imediat spre dreapta sau spre stînga poziției lor inițiale și se mută doar o dată. Această implică faptul că, indiferent de ghișeele disponibile, timpul de așteptare al călătorilor nu este mai bun decît cel dat de grupuri de trei ghișee cu acces limitat.

În fig. 10.13 se interpretează aceste fapte; ea a fost construită după cum urmează:

1. Orele de aglomerație pentru fiecare din cele cinci zile au fost împărțite în subintervale de 15 min.

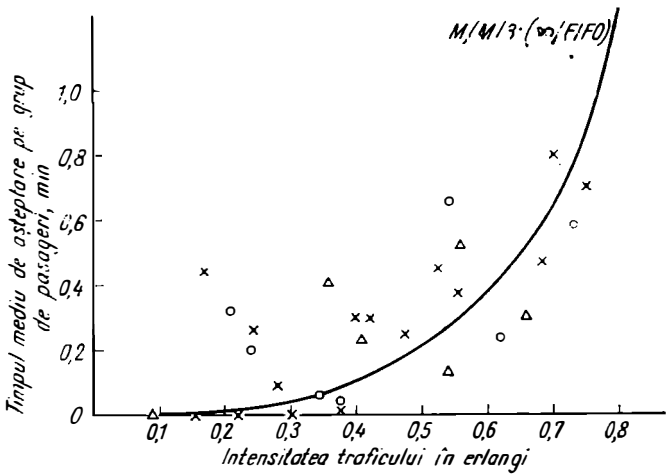


Fig. 10.13. Timpul mediu de așteptare și intensitatea traficului (la ghișeele obișnuite.)

Δ patru ghișee. × cinci ghișee. o șase ghișee.

2. A fost înregistrat fluxul total care sosește și este prelucrat la fiecare ghișeu și, de asemenea, timpul de așteptare al pasagerilor și timpul de relaxare al tuturor angajaților.

3. Au fost calculați timpul mediu de așteptare al pasagerilor și utilizarea angajaților pentru fiecare interval de timp.

4. Rezultatele au fost reprezentate grafic pentru acele subintervale în care traficul este aproximativ egal.

Este evident că timpii de așteptare pentru ghișeele 4, 5 și 6 deschise sînt mai mari decît cei obținuți teoretic din modelul $M/M/c$:(∞/ /FIFO) datorită solicitării diferite a ghișeelor (ghișeele erau puse în funcțiune de obicei în ordinea 6,7,5,8,4,3,2). Dar timpii de așteptare pentru ghișeele 5 și 6 deschise, în loc să fie mai scăzuți decît cel pentru ghișeul 4 deschise la orice nivel de utilizare, erau chiar mai ridicați.

Membrii echipei au examinat amănunţit repartiţia timpului de servire, ea a fost găsită aproximativ identică cu cea observată la ghişeele de verificare de zbor cu doi ani în urmă; ambele sînt comparate în fig. 10.14. Aceasta este neliniştitor căci la un ghişeu obişnuit

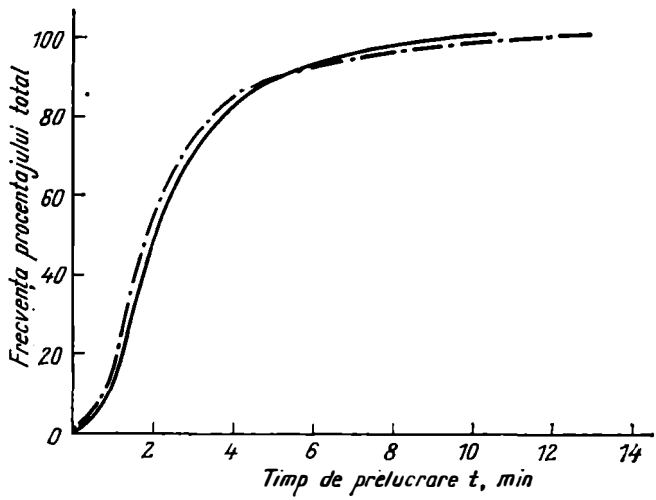


Fig. 10.14. Comparaţie între repartiţiile timpului de servire.
---- Observaţii în 1960, ghişee de zbor.
— · — Observaţii în 1958, ghişee comune

de verificare sînt mai puţine de făcut decît la un ghişeu de zbor. Care ar putea fi motivele? Cercetătorii au strîns informaţii referitoare la timpul mediu de servire şi la volumul de pasageri pentru fiecare interval de 30 min din mai multe zile şi le-au corelat (v. fig. 10.15). Timpul mediu de servire este corelat cu volumul de lucru. Acelaşi rezultat a fost obţinut şi pentru cazul în care funcţionează ghişeele 4 şi 6. Deci angajaţii lucrează mai repede atunci cînd văd mulţi pasageri în faţa lor. În timpul orelor de aglomeraţie timpul mediu de servire a unui pasager a fost de numai 1,50 min, mai mare decît 1,13 min planificate. Cozile din perioadele de vîrf pot fi atribuite parţial unui timp de servire nedorit de lung. (Dar cititorul să noteze legea lui Parkinson : „munca se extinde pentru a umple timpul dis-

ponibil pentru îndeplinirea ei".) De ce chiar în timpul perioadelor aglomerate timpul mediu de servire trebuie să fie mai lung decît cel prevăzut? Studiile rezervării de locuri pentru transportul aerian care au fost întreprinse în intervalul dintre a doua serie de studii de verificare și aplicarea în practică a rezultatelor lor a sugerat răspunsul. Un angajat de la un ghișeu de verificare de zbor trebuie să rețină informații despre un singur zbor: la ghișeul obișnuit trebuie să-și rețină informații despre 12 sau 16 sau chiar 20 de zboruri simultan. Membrii echipei au mai aflat că aproximativ 40% din pasageri (procentul variază de la zbor la zbor) cer informații în timpul verificării; mai mult, ei au observat că angajații trebuie să consulte orarul zilnic pentru a da răspunsuri. Acest lucru sugerează că solicitarea

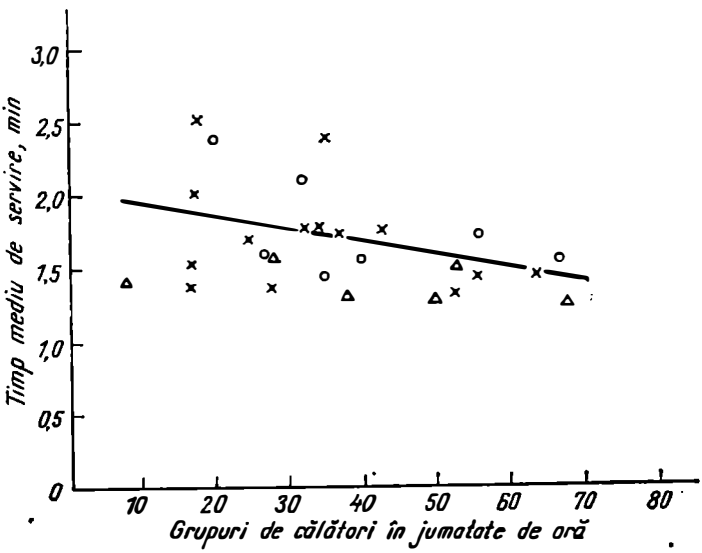


Fig. 10.15. Timpul de servire și volumul de muncă.
Δ patru ghișee. × cinci ghișee. o șase ghișee.

mărită a memoriei funcționarilor poate să fie cauza unui timp de servire mai lung. Nu s-a constatat diferențe semnificative între timpii medii de servire ai diferiților angajați, cu excepția a două tinere care au venit în secție abia de două săptămîni. Ambele erau mai

încete decît media. Deci singura explicaţie raţională pare să fie solicitarea mai mare a memoriei angajaţilor şi comportarea ciudată a pasagerilor în aşteptare.

Ca rezultat, s-a constatat că pentru ca verificarea să fie eficientă, clădirea în care se află secţia trebuia să fie modificată în bună parte. De exemplu, uşile de intrare spre spaţiul de verificare trebuiau să fie mărite pentru a se evita concentrarea călătorilor la unele ghişee. Angajaţii trebuiau să aibă mijloace de acces mai rapid la informaţiile referitoare la zbor, cum ar fi televiziunea în circuit închis. Aici se încheie misiunea cercetătorilor, este rîndul arhitecţilor şi constructorilor. Suprafaţa reconstruită a fost dată în funcţiune la începutul anului 1961.

10.4. COMENTARIU

O atenţie deosebită trebuie acordată cazurilor, cînd rezultatele modelelor clasice de aşteptare se aplică unor sisteme în care atît staţiile de servire cît şi clienţii sînt oameni.

Două dintre cele mai comune ipoteze, pe care se bazează modelele obişnuite, sînt adesea invalidate; anume:

1. Repartiţia timpului de servire nu este neafectată de sistemul de intrare. Adesea timpul mediu de servire la un moment t este corelat cu intervalul mediu între sosiri la un anumit moment anterior $t-\tau$ (τ poate fi foarte scurt sau chiar nul dacă staţia de servire-om este capabilă să detecteze rapid schimbările ce au loc în intensitatea intrărilor).

2. Dacă sînt mai mult de două staţii de servire, regulile FIFO nu se respectă; în general ele se respectă doar dacă se forţează acest lucru.

Totuşi analistul nu trebuie să renunţe la verificarea valabilităţii acestor presupuneri atunci cînd studiază sisteme ca cel prezentat în acest capitol.

11

PROBLEMA SORTĂRII CUPOANELOR

Vom examina în cele ce urmează o situaţie practică în care apare o singură staţie de servire. În fig. 11.1 este schiţată amplasarea ghişeelor comune de verificare a documentelor la un aeroport. Pasagerii se aşază în aşteptare în faţa unui număr de ghişee. La sosirea la ghişeu pasagerul îşi prezintă biletul, i se cîntăreşte bagajul, greutatea bagajului se înscrie pe bilet, apoi se detaşează cuponul. Ne va interesa ce se întîmplă în continuare cu acest cupon.

Funcţionarul pune cuponul pasagerilor pe un transportor de documente, care-l transportă spre punctul de sortare. Aici cuponul

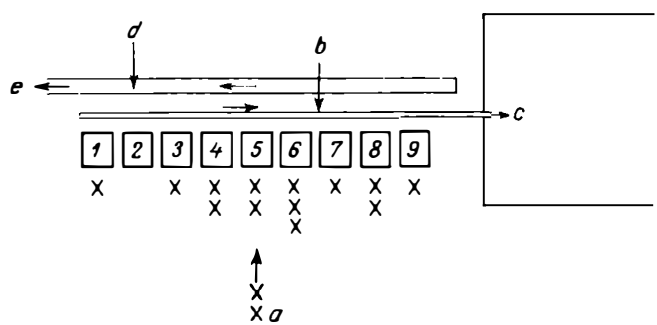


Fig. 11.1. Schiţa amplasării transportoarelor de bagaje.
a — pasageri; b — transportor documente; c — control şi sortare a cupoanelor; d — transportor bagaje; e — spre punct de sortare.

cade într-o cutie. Un funcţionar ia cupoanele din cutie unul cîte unul, le verifică şi le repartizează în casete, după indicativul zborului înscris pe fiecare cupon. Întîrzierile survenite în sortarea cupoanelor

pot produce întîrzieri în plecarea avioanelor şi de aceea este foarte important ca sortarea să fie rapidă. Dar acest lucru este greu de realizat prin simpla dublare a celor care sortează, deoarece, fără să vrea, aceştia interferează şi se stînjenesc reciproc.

Din acest motiv, prezintă interes determinarea capacităţii maxime de sortare a unui singur funcţionar. Aceasta poate limita numărul ghişeelor care alimentează un singur punct de sortare, ceea ce are consecinţe asupra proiectării spaţiilor de verificare a călătorilor, consecinţe care nu pot fi ignorate. Studiile ergonomice au condus la un proiect excelent al locului de sortare şi nu se întrevede nici o perspectivă pentru o reducere ulterioară a timpului necesitat pentru sortarea fiecărui cupon. Care este capacitatea unui funcţionar care sortează singur? Sau, mai general, care este repartiţia timpilor de aşteptare, la diferite intensităţi ale traficului în sistemul de aşteptare cu o singură staţie de servire în care: „clienţii" sînt cupoanele, „momentul sosirii" este momentul căderii cupoanelor în cutie, iar „staţia de servire" este funcţionarul care sortează. Ce putem să spunem despre intrările în sistem? Numărul cupoanelor care sosesc în punctul de sortare (să zicem într-o zi) este egal cu numărul pasagerilor care iau loc în avioanele ce au plecările în acelaşi interval de timp (ne permitem să ignorăm abaterea). Este clar că numărul pasagerilor verificaţi într-o oră este legat de numărul zborurilor ce au loc într-o anumită perioadă de timp. De asemenea este bine cunoscut faptul că zborurile nu sînt programate uniform în tot timpul zilei. Prin urmare, intensitatea medie a cupoanelor sosite în cutie nu este constantă. Cititorul familiarizat cu ideea de „perioadă ocupată" ne spune că aceasta nu are nimic de-a face cu intrările considerate în problema determinării capacităţii maxime. În lucrul la capacitate maximă intensitatea medie a sosirilor cupoanelor poate fi considerată constantă pentru o perioadă mai lungă. Aşa şi este, dar apar două efecte secundare ale modificării intensităţii intrărilor, efecte cu o anumită importanţă practică.

Întîi, oare timpul mediu de sortare al unui cupon de către un funcţionar variază odată cu variaţia intensităţii sosirii cupoanelor? Dacă nu, probabil că totul este în ordine. Dar dacă da, oare în ce fel variază şi ce se întîmplă? Există oare semne de stabilizare ce apar odată cu creşterea intensităţii de sosire? Dacă este aşa, atunci poate am putea calcula timpul de servire care corespunde intensităţii maxime de sosire. Dar dacă n-ar apărea semne de stabilizare? Observaţiile au arătat că fără îndoială funcţionarii care se ocupă de

sosirea lucrează mai repede cînd au teancuri de cupoane nesortate decît atunci cînd sosirea cupoanelor are loc mai rar. Dar s-a constatat de asemenea că pe măsură ce creşte intensitatea sosirilor, funcţionarii tind să ajungă la un timp mediu minim de sortare şi nu mai pot să-l îmbunătăţească. Se impune o restricţie mai serioasă asupra timpului mediu de servire: limitarea proporţiei cupoanelor sortate greşit.

Unul dintre supraveghetorii de la un aeroport unde se studiau aceste probleme a sugerat că proporţia sortărilor greşite creşte în timpul perioadelor de aglomeraţie. S-au făcut cercetări şi s-a degajat cu claritate un prag peste care greşelile deveneau frecvente. Ca rezultat al acestor cercetări, echipa de cercetători a stabilit că timpul de sortare minim al unui cupon, timp care poate fi folosit în calcule, este de 2,1 s cu un coeficient de variaţie mai mic de 0,1. De fapt repartiţia timpului de servire a fost găsită constantă.

Să revenim la intrările în sistem. O analiză la întîmplare a momentelor sosirii în timpul unor perioade aglomerate scurte a arătat că repartiţia timpilor dintre sosiri este exponenţial negativă. Totuşi trebuie să privim acest rezultat cu grijă: sosirea cupoanelor în punctul de sortare a fost generată de plecările pasagerilor de la un număr finit şi, de fapt mic, de ghişee de verificare, număr care nu depăşea 11 în cele mai aglomerate perioade. În perioadele de aglomeraţie întotdeauna este cineva care aşteaptă sau este servit la fiecare ghişeu de verificare. Plecările de la fiecare ghişeu au loc în momente separate de un timp de servire. Aşadar, la fiecare ghişeu intervalul dintre plecări are o repartiţie identică cu repartiţia timpului de servire. Acestea sînt de tip Euler. Dacă nu am lua în consideraţie diferenţele de timp (mici) între plecările pasagerilor de la diferitele ghişee şi sosirile cupoanelor în cutie, datorite drumului mai lung sau mai scurt pe banda transportoare a cupoanelor, atunci timpul dintre sosiri în punctul de sortare are aceeaşi repartiţie ca şi ieşirile cumulate ale mai multor procese de reînnoire cu repartiţii identice.

Cox şi Smith au arătat că în aceste condiţii repartiţia ieşirilor cumulate este aproximativ exponenţială negativă. A fost stabilită astfel o motivaţie pentru rezultatul obţinut empiric. Acest argument nu este deloc valabil pentru perioadele de relaxare, însă pe noi nu ne interesează decît perioadele de vîrf. Aşadar, pînă acum avem (aproximativ) sosiri Poisson şi un timp de servire constant. Modelul în codul Kendall este $M/D/1$. Totuşi modelul nu cuprinde încă disciplina de aşteptare. Cupoanele care vin pe transportor cad în cutie deasupra

sosirilor anterioare sau plutesc şi se amestecă printre celelalte cupoane. Dacă funcţionarul care sortează este mai comod, el ia cupoanele în ordinea în care le găseşte şi astfel ordinea de servire este apropiată de regula LIFO. Dar nu este întotdeauna aşa. Unii funcţionari scot mai întîi cupoanele de dedesubt şi astfel ordinea de servire este apropiată de regula FIFO. Însă cutia este destul de mare faţă de dimensiunile cupoanelor, care pot să rămînă pe pereţii cutiei în loc să cadă la fund. Toţi aceşti factori fac ca disciplina de aşteptare să semene foarte mult cu regula SIRO care deşi nu este foarte corectă pentru descrierea procesului, este mult mai realistă decît regulile FIFO sau LIFO.

Aşadar, procesul de aşteptare poate fi reprezentat prin modelul $M/D/1:(\infty/\text{SIRO})$ fapt care permite să se prevadă probabilitatea ca un cupon să întîrzie în cutie în timp mai mult sau mai puţin îndelungat. Întîrzierea admisibilă este destul de mică : maximum 10 % din cupoane pot să întîrzie mai mult de 10 s.

În anul cînd s-a studiat această problemă tabelele lui Burke pentru modelul $M/D/1:(\infty/\text{SIRO})$ nu erau disponibile ca să uşureze munca. Echipa de cercetători nu avea tabele decît pentru : modelele $M/M/1:(\infty/\text{SIRO})$, $M/M/1:(\infty/\text{FIFO})$, $M/D/1:(\infty/\text{FIFO})$ şi nu prea mult timp la dispoziţie (căci erau în studiu şi probleme mai importante). Dacă aceste trei modele sînt considerate în ordinea în care, pentru orice valoare a traficului, produc cei mai puţini timpi lungi de aşteptare, atunci trebuie să le scriem astfel : $M/D/1:(\infty/\text{FIFO})$, $M/M/1:(\infty/\text{FIFO})$, $M/M/1:(\infty/\text{SIRO})$. Este logic să se presupună că modelul care ne interesează, anume $M/D/1:(\infty/\text{SIRO})$ produce mai mulţi timpi de aşteptare lungi decît primul, dar mai puţini decît ultimul. Or, pentru un singur funcţionar, utilizarea maximă acceptabilă ce menţine timpul de aşteptare în limitele prescrise se încadrează între : 80 % din cea corespunzătoare primului model şi 53 % din cea corespunzătoare celui de-al treilea model. (Ambele procente au fost citite pe grafice şi se poate să se fi strecurat şi greşeli de citire.)

Intensitatea maximă de intrare a cupoanelor trebuie deci să situeze între circa 910 pînă la aproximativ 1370 cupoane pe oră. În plus soluţia poate fi menţinută cel puţin cinci ani.

Echipa de cercetători a găsit o soluţie operativă fără a rezolva în sens matematic problema formală. Ulterior a apărut articolul lui Burke (v. cap. 3). Utilizarea maximă dată de modelul $M/D/1:(\infty/$ /SIRO) este, după Burke, de aproximativ 79 %. Cu alte cuvinte, nu este o diferenţă prea mare între cele două modele $M/D/1:(\infty/$FIRO) şi $M/D/1:(\infty/$SIRO). Pentru noi a fost o satisfacţie să aflăm că aprecierea pe care am făcut-o era atît de aproape de cea corectă şi că soluţia pe care am dat-o a fost valabilă fără modificări cel puţin trei ani.

12

PATRU SITUAȚII SELECTATE

În acest capitol vom considera pe scurt patru situații în care a fost aplicată teoria așteptării pentru a se obține soluții operaționale ale unor probleme reale.

12.1. PROIECTAREA REȚELEI DE PISTE ÎN AEROPORT

Vom analiza investigațiile făcute în Anglia de către matematicienii Laboratorului național de fizică în colaborare cu Ministerul Aviației Civile. S-au făcut cercetări pentru proiectarea sistemului de căi de rulare și de acces pentru noul aeroport din Londra, care era în curs de construcție prin 1950. N-aș putea prezenta mai bine problema și rezultatele obținute decît unul dintre cei ce au participat efectiv, anume K. D. Tocher, ale cărui spuse le reproduc cuvînt cu cuvînt:

„Fluxul de material printr-un sistem de transport este limitat de capacitatea sistemului, de posibilitățile de întreținere și păstrare; ele determină în unele locuri gîtuiri care încetinesc întregul proces. De obicei în astfel de situații este chemat să îmbunătățească situația un departament de cercetări operaționale. Se pune problema dacă este posibil să se proiecteze un sistem care să reducă pe cît posibil acea încetinire a procesului care este datorată unor variații aleatoare. Teoria așteptării rezolvă rar aceste probleme în întregime, deoarece este o teorie descriptivă care prezice doar efectul variației în sistem.

Odată ce disciplina de așteptare depinde de starea întregului sistem, modelele matematice dezvoltate de către teoreticieni nu mai sînt folositoare. Devine necesară o teorie a servomecanismelor aleatoare; însă pînă în prezent nu au fost întreprinse studii serioase în acest domeniu. Totuși teoria așteptării poate să dea adesea indicații folositoare în cazuri de probleme de congestie a unui sistem. Ea a fost aplicată mai de mult în proiectarea unui aeroport. S-a propus ca aeroportul să asigure deplasări libere între pistele de rulare și/sau platformele de încărcare și de descărcare și/sau hangare, cu singura condiție ca fiecare avion să urmeze un traseu fix. Au fost propuse mai multe sisteme de îndrumare și se cerea să se determine acel sistem care asigură debitul maxim de avioane, precum și acest debit maxim".

Să facem două observații: prima, că Tocher și colegii săi au abordat două dintre tipurile de probleme operaționale menționate în capitolele 6 și 7, anume comparația între sisteme diferite și estimarea unor parametri ai unui sistem. Ne așteptăm, în lumina celor cunoscute, ca al doilea tip să fie mult mai dificil. Tocher continuă:

„Fiecare sistem de îndrumare conținea intersecții la care avionul ar fi putut fi oprit. Aceasta era una dintre sursele de congestie. Fiecare platformă avea o capacitate limitată și s-a pus restricția ca doar un avion să se poată deplasa spre o platformă liberă. Astfel sistemul poate fi modelat ca rețea de șiruri de așteptare alimentată inițial de mai multe piste de rulare. Dacă luăm în considerare modelul matematic general, această rețea conduce la o problemă matematică imensă. Totuși, dacă presupunem sistemul statistic stabil și dacă considerăm doar starea de echilibru, se pot face oarecari progrese.

Vom presupune că timpul necesar pentru traversarea unei intersecții, timpul de încărcare și de descărcare au repartiții exponențiale negative. În această ipoteză problema se separă într-o mulțime de șiruri de așteptare independente. Pentru fiecare astfel de șir de așteptare, singurele date ce ne interesează sînt intensitatea medie a intrărilor și timpul mediu de oprire (blocare). Odată stabilite regulile de circulație și densitatea traficului pe diferitele drumuri, intensitățile de intrare se obțin ca soluție a unui sistem de ecuații liniare. Astfel, devine posibil să se evalueze mărimea șirului de așteptare în fiecare punct de aglomerație și astfel să se prezinte o vedere de ansamblu a numărului de avioane care așteaptă în diferitele puncte ale sistemului în orice moment. Din punctul de vedere al pasagerului prezintă interes doar timpul total de așteptare și numărul de opriri interme-

diare. Și acestea pot fi deduse folosind modelul, dar cu prețul unor calcule ceva mai complicate.

Ceea ce se obține astfel nu poate fi acceptat în mod absolut, dar le putem folosi pentru studii comparative pentru a elimina unele dintre sistemele propuse.

Stabilirea nivelului maxim al debitului este foarte simplă. Fiecare șir de așteptare individual trebuie să fie stabil și astfel problema se reduce la studierea densității traficului care saturează diversele șiruri de așteptare. Astfel apar cu claritate și intersecțile care produc gîtuiri ale circulației.

A mai fost posibil să se stabilească, pentru un sistem de îndrumare dat, cea mai bună dirijare a avioanelor pentru diverse situații. S-a putut stabili astfel că unul dintre sistemele de îndrumare era suficient de bun (pentru densitatea de trafic estimată). Întregul studiu a relevat posibilitatea obținerii unui flux mare, și în final s-a renunțat la îndrumarea arbitrară în favoarea uneia strict controlate."

Această problemă a ridicat și chestiunea alegerii după un criteriu bazat pe timpii de așteptare sau pe timpii totali de așteptare. Dacă, să zicem, trei sisteme conduc toate la un flux acceptabil de mare, atunci va trebui ales între ele pe cel care asigură cel mai scăzut timp de așteptare, sau pe cel care satisface anumite condiții asupra timpilor de așteptare și este cel mai ieftin.

12.2. PROBLEMA AUTOBUZELOR ÎN AEROPORT

În aeroportul din Londra avioanele care vin sau urmează să decoleze se află în spații care nu sînt în imediata apropiere a clădirilor aeroportului. De cele mai multe ori sînt necesare autobuze pentru transportul pasagerilor de la clădire la avion sau invers. Una din metodele folosite impune existența în rezervă a unui grup de autobuze pe o platformă, care să asigure transportul ori de cîte ori este nevoie. Numărul autobuzelor chemate variază, depășind cîteodată numărul celor existente, deoarece sosirile și plecările avioanelor variază din diverse motive în jurul momentelor planificate. Cercetătorii au fost interesați să calculeze numărul minim de autobuze care trebuie ținute în rezervă pe platformă în diferitele momente ale zilei pentru a fi siguri că cel mult un procent mic de solicitări nu pot fi onorate

imediat. Această problemă poate fi considerată o problemă de aşteptare. Autobuzele sînt staţii de servire, avioanele care pleacă (sau sosesc) sînt clienţii, iar timpii de servire sînt aceia în care autobuzele sînt ocupate cu transportul pasagerilor către sau de la avioane. Aceşti timpi sînt de fapt timpi de blocare, adică timpul de servire este acel timp pentru care un avion are nevoie de un autobuz. Timpul de servire a fost găsit aproximativ constant cu o medie de 7,5 minute şi o abatere standard de 1,7 minute.

Intrările în sistemul de aşteptare sînt totuşi departe de a fi repartizate Poisson. Analiza datelor înregistrate a arătat că intrările într-o perioadă de timp de o oră sînt repartizate aproape Poisson, cu o medie dată de

$$m_r = \frac{1}{4}(s_{r-1} + 2s_r + s_{r+1}),$$

unde s_r reprezintă numărul zborurilor planificate în ora r, iar a_r numărul zborurilor reale.

Aşadar, în timpul oricărei ore, sistemul pare bine descris de modelul $M/D/c :(\infty/\text{FIFO})$. Din păcate, numărul pasagerilor care urmau să fie transportaţi este variabil de la zbor la zbor, astfel încît procentul π_1 de zboruri necesită un singur autobuz, în timp ce procentul π_2 necesită două ($\pi_1 + \pi_2 = 1$). Aşadar, intrările constau de fapt în două intrări de tip Poisson; una pentru „clienţi" singuri, avînd intensitatea $\pi_1 m_r$ şi alta pentru perechi de clienţi, avînd intensitatea $\pi_2 m_r$. Modelul care rezultă, $M(\pi_1 m,\ \pi_2 m)/D/c :(\infty/\text{FIFO})$, este total descurajator şi foarte greu i se pot găsi soluţiile analitice.

Cercetătorii au imaginat un procedeu de calcul pe care l-au folosit pentru a calcula valorile probabilităţii existenţei (în echilibru statistic) a cel mult i clienţi simultan. Apoi cercetătorii au întocmit grafice legînd : a) procentajul zborurilor servite necorespunzător (prin lipsa unui autobuz sau chiar a două), din : b) numărul zborurilor pe oră, pentru : c) două pînă la zece autobuze existente în rezervă. Problema a fost astfel rezolvată.

Interesul acestui studiu constă în faptul că este o ilustrare a modului în care cercetătorii au prezis intensitatea intrărilor cu ajutorul unei formule de interpolare, obţinînd astfel că intrările sînt aproximativ poissoniene pentru perioade de o oră.

12.3. ÎNTÎRZIERI ALE CIRCULAȚIEI ÎN PUNCTELE DE VAMĂ

Unul din cele mai renumite studii asupra problemelor reale de așteptare a fost întreprins în 1950. Acest studiu a cucerit premiul Lanchester. Aici se prezintă sumar rezultatele muncii lui L. C. Edie, autorul studiului.

Una dintre activitățile cheie ale personalului de poliție folosit de către Administrația Portului New York este colectarea taxelor la poduri și la tunele. În 1952, de exemplu, rulajul financiar generat de această sarcină depășea 1 000 000 $ pe an. Analizele preliminare ale operațiilor de la punctele de vamă au relevat faptul că aceste cheltuieli nu aveau deloc efectul optim : standardul de servire era foarte variabil de la un punct la altul și de la un moment la altul ; de asemenea folosirea ofițerilor de trafic era foarte neregulată.

Se punea problema evaluării standardului de servire oferit clienților la punctele de vamă, și a evaluării modului în care variază acest standard odată cu variația volumului de trafic ; de asemenea, se punea problema stabilirii standardului de servire ce trebuie oferit și în final de a găsi modalități de a planifica personalul în modul cel mai economic pentru a obține acest standard.

S-au cheltuit mai multe sute de om/ore de muncă pentru a obține datele empirice, pe baza cărora să se decidă modelul cel mai convenabil de folosit în studierea sistemului. După o muncă ulterioară la fel de lungă pentru analizarea acestor date, Edie și grupul său de cercetători au ajuns să obțină un model satisfăcător. Detaliile pot fi găsite în articolul lui Edie* ; nu avem spațiul necesar pentru a le reproduce.

Intrările aveau o repartiție Poisson la volume de trafic de pînă la circa 600 vehicule/oră și o repartiție normală peste această valoare. Cercetătorii au tras concluzia că acest efect corespunde apropierii repartițiilor Poisson și normală de repartiția binomială și au considerat apoi că numărul sosirilor pe unitatea de timp (luată a fi egală cu 30 s) este repartizat binomial.

Repartiția timpului de servire a fost găsită constantă. Acest rezultat s-a stabilit pe două căi : mai întîi prin măsurarea unor timpi de servire, apoi prin comparația între datele observate la un punct de

* Edie, L. C. *Trafic Delays at Toll Booths.* În : Oper. Research, **2,** nr. 2, 1954.

vamă privind legătura între timpii de așteptare și intensitatea traficului și între datele teoretice obținute din modelele $M/M/1$:(∞/FIFO) și $M/D/1$:(∞/FIFO). Însă atunci cînd cercetătorii au analizat situația pentru mai multe puncte de vamă, au observat că datele observate diferă mai mult de datele teoretice obținute pe baza modelelor $M/M/c$: (∞/FIFO) și $M/D/c$: (∞/FIFO). Motivul era că, în cazul mai multor stații de servire, clienții nu respectau strict regula FIFO. Destul de frecvent un vehicul se așeza într-un șir de așteptare lung, în timp ce pe altă linie nu erau deloc vehicule în așteptare. Aceste observații au părut suficiente că să explice discrepanța dintre rezultatele teoretice și cele observate.

Edie a mai descoperit faptul că timpii de servire sînt legați de volumul de trafic. El a folosit relațiile pe care le-a stabilit între intensitatea traficului și timpii de așteptare pe de o parte, și timpii de servire și volumul de trafic pe de altă parte, pentru a determina dependența între timpul mediu de așteptare al unui vehicul și volumul de trafic (în vehicule pe zi).

Următoarea fază a studiului a fost determinarea repartiției lungimilor șirurilor de așteptare formate în fața punctelor de vamă. Și aceasta a fost făcută prin metode asemănătoare cu cele folosite în faza precedentă. În această fază Edie a considerat important să identifice cît de mult timp pot să întîrzie șoferii într-un șir de așteptare pînă la a încerca posibilitățile de deplasare.

Următoarea fază, căreia din nefericire Edie i-a acordat mai puțină atenție în articolul său, a constat în prognoza volumului orar de trafic în fiecare zi a săptămînii. Modul de variație a volumului de trafic a fost găsit identic pentru zilele de marți, joi și vineri ; celelalte zile au necesitat un tratament individual. S-a găsit că împrăștierea între traficul mediu și cel de vîrf era între 10 și 60 procente la poduri și între 10 și 30 procente la tunele. Bineînțeles că în aceste împrejurări nu poate fi posibil să previzionăm cu o precizie mare traficul într-un punct particular de vamă, la o anumită oră dintr-o anumită zi a săptămînii.

Cititorul se poate mira ce volum mare de date a trebuit prelucrat pentru a obține un model precis al sistemului de așteptare.

În final, pe baza timpului mediu de așteptare determinat în mod optim pentru un volum de trafic mediu (prognozat), s-a calculat numărul de puncte de vamă necesare să fie deschise, din jumătate în jumătate de oră. Apoi acestea au fost ajustate pentru a reduce traficul de vîrf la un nivel acceptabil. Astfel, s-a obținut o planificare

săptămînală a punctelor de vamă necesare și a rămas doar problema planificării personalului (despre care însă Edie nu ne mai dă amănunte în lucrarea sa).

12.4. SISTEME DE ÎNTÎLNIRE

Un tip de sisteme de așteptare care nu a fost discutat în capitolele teoretice din această carte, ci doar amintit, este cel în care momentele de sosire a clienților sînt dinainte planificate. Două exemple : sosirile avioanelor la un aeroport (problemă de care se leagă al doilea exemplu din acest capitol) și sosirea navelor cu minereu la cheiul unei uzine siderurgice. În ambele cazuri momentele reale de sosire diferă de cele planificate și, după cum am observat deja cu cîteva ocazii, de obicei, este posibil să aproximăm modul de sosire real printr-unul poissonian. Dar ce se întîmplă dacă clienții sosesc la momentele planificate? Acest tip de sistem a fost studiat în legătură cu sistemele de programare a solicitanților la policlinici, iar sistemele de acest tip sînt numite de întîlnire.

Principalele modele studiate teoretic presupun că duratele dintre sosiri, atît planificate cît și reale, sînt constante, astfel încît intrările sînt regulate. Însă abordarea analitică a acestor sisteme devine destul de dificilă, deoarece ele se abat de la presupunerile simple asupra repartiției timpilor de servire. Cu toate că modelele ca de exemplu $D/M/1 :(\infty/\text{FIFO})$ sînt interesante, ele nu sînt prea folositoare.

Mai recent, au fost publicate trei rapoarte care descriu cum au fost studiate unele sisteme de întîlnire. Unul dintre acestea, datorat lui Fry, prezintă un deosebit interes pentru că descrie experimentele făcute asupra unui sistem de așteptare, mai curînd decît mînuirea unui model. Altul, datorat lui Welch, se ocupă de unele aspecte ale comportării iresponsabile a stațiilor de servire (în cazul nostru medici), care nu încep munca la timp, și de efectele acestei comportări asupra sistemului de așteptare. Și acesta este extrem de interesant, deoarece există situații analoage în industrie care se pare că n-au fost încă discutate nicăieri. De exemplu, avioanele planificate să sosească la un aeroport găsesc pistele de rulare ocupate și nu pot ateriza. Ultimul articol îl vom prezenta mai în detaliu.

Autorul articolului, R.R.P. Jackson, este un proeminent teoretician al aşteptării. Problema abordată de el constă în conceperea unui sistem de întîlnire eficient pentru cadrele medicale (trebuie să spunem că majoritatea medicilor britanici nu-şi consultă pacienţii în cabinet pe bază de programare şi că există o altă problemă, mai fundamentală decît cea a determinării sistemului optim de întîlnire — anume convingerea medicilor că trebuie să adopte un sistem de întîlnire; în schimb, în Canada şi S.U.A. este o regulă să se planifice întîlniri între medici şi pacienţi).

Jackson abordează un număr de probleme practice ce apar în legătură cu sistemele de întîlnire pentru cadrele medicale, de exemplu apariţia urgenţelor. El arată că modelul de aşteptare adecvat situaţiei este $D/E/1 :(\infty/\text{FIFO})$. De fapt, presupunerea că sistemul are capacitatea infinită nu este explicit formulată şi, dacă ar fi o cameră de aşteptare limitată, atunci modelul adecvat ar fi $D/E_k/1 :(L/\text{FIFO})$. În orice caz, metodele analitice existente sînt nepotrivite cu scopul urmărit, iar Jackson decide să simuleze modelul cu ajutorul unui calculator electronic (şi folosind un limbaj special de simulare). Această decizie, luată de un cercetător experimentat, este ilustrativă pentru limitele abordării teoretice (analitice) a problemelor de aşteptare.

Deoarece fiecare program de consultaţii era limitat la două ore (şi ne imaginăm că la sfîrşitul său doctorul amînă toţi pacienţii care încă se află în aşteptare), Jackson alege ca măsură pentru eficienţa doctorului timpul total pierdut de pacienţi aşteptînd. Cu ajutorul simulării Jackson obţine estimări ale timpului mediu de aşteptare al unui pacient, pentru diverse combinaţii ale timpului mediu de consultaţie şi ale duratelor medii între sosirile pacienţilor. Prin tabelarea lor se obţine intervalul optim între sosiri.

Jackson a cercetat şi o soluţie aproximativă, comparînd rezultatele simulării pentru modelul $D/E_k/1 :(\infty/\text{FIFO})$ cu cele obţinute din formulele pentru modelul $E_k/M/1 :(\infty/\text{FIFO})$, pe care le-a stabilit anterior împreună cu Nickols. Ca rezultat al acestor studii, Jackson a conceput un sistem optim de programare a pacienţilor, care a fost apoi folosit de către însuşi medicul lui Jackson, cu avantaje pentru ambele părţi.

13

PROIECTAREA SEDIULUI DIN LONDRA AL COMPANIEI BEA

13.1. PREZENTAREA PROBLEMEI

Multe din problemele ridicate în cadrul acestei operaţii sînt legate de unele sisteme de aşteptare. Operaţiile de verificare a pasagerilor prezentate în cap. 10 sînt edificatoare în acest sens. Evident că mai există şi multe alte exemple. În cadrul acestui capitol vom prezenta trei dintre acestea. Fiecare din ele este un exemplu de alt fel de sistem, dar datorită faptului că apar în situaţii operaţionale apropiate, este necesar a le trata împreună.

Companiile aeriene care organizează zboruri între oraşe mari, cum ar fi New York sau Londra, au considerat eficient a crea sedii în oraş pentru pasagerii lor. Să analizăm de exemplu cazul Londrei, în anul 1958. În acea perioadă majoritatea pasagerilor de la Londra — lucru care poate este valabil şi astăzi — puteau fi repartizaţi în una din următoarele trei clase :

1. Pasagerii care locuiau în Londra şi care-şi începeau călătoria în Londra. Majoritatea acestora proveneau din districtele din zona central-vestică a Londrei, precum şi din suburbiile vestice.

2. Pasagerii originari din alte zone ale Marii Britanii, care au călătorit spre Londra cu trenul sau autobuzul, pentru a lua avionul de la Londra. Principalele staţii feroviare şi de autobuz se află în districtele vestice sau central-vestice ale oraşului, dar în nici un caz lîngă aeroportul principal, Heathrow.

3. Pasagerii din alte ţări, care au vizitat Londra şi au locuit în hotelurile londoneze. Acestea se află în districtele vestice sau central-vestice ale oraşului.

Clasificări similare se pot face — şi chiar au fost făcute — pentru pasagerii ce călătoresc spre Anglia, via Londra, venind din alte ţări.

Este limpede că un sediu localizat în zona vestică a oraşului va fi un loc convenabil pentru a servi ca punct de colectare pentru pasagerii de la Londra şi ca un punct de destinaţie pentru pasagerii spre Londra. Există anumite avantaje atît din punctul de vedere al călătorului, cît şi din cel al companiei aeriene. Mai întîi, operaţia de transport al pasagerilor cu autobuze între sediul din oraş şi aeroport constituie un mijloc mai ieftin pentru pasageri. De exemplu, pentru două persoane, costul taxiului comparativ cu costul biletului de autobuz (cursa care duce la aeroport) este de 3 : 2 la Montreal, 2 : 1 la New York şi 3 : 1 la Londra. În al doilea rînd, pasagerul simte o anumită siguranţă, deoarece odată verificat la sediul din oraş, el nu mai are probleme legate de rezervarea locului, de bagaje sau de posibilitatea de a pierde cursa datorită întîrzierii. În sfîrşit, o companie aerieană consideră ca foarte eficientă cunoaşterea numărului de pasageri ce vor utiliza o anumită cursă, cu mai mult timp înainte de zbor. Avantajele pe care le prezintă un sediu în oraş, atît pentru pasageri, cît şi pentru companiile aeriene, variază de la oraş la oraş, iar global au tendinţă de diminuare odată cu diminuarea mărimii oraşului respectiv.

În primăvara lui 1958, conducerea Companiei BEA, conştientă de continua mărire a traficului pasagerilor, a decis să construiască un nou sediu în Londra şi să înlocuiască complet vechiul sistem de verificare a pasagerilor. Din motivele expuse anterior, noul sediu va trebui construit într-unul din cartierele vestice ale oraşului. S-a decis că cea mai bună amplasare, din cele posibile, era chiar locul ocupat de vechiul sediu.

Deşi, din punct de vedere strategic, aceasta era o decizie excelentă, ea a pus o serie de probleme dificile, datorită faptului că pămîntul era stîncos, existau deja construcţii, iar dedesubt trecea o linie de metro de maximă importanţă. Spaţiul era extrem de limitat şi era limpede că cea mai mare problemă era aceea de a folosi cu maximum de eficienţă spaţiul existent. Direcţiunea companiei a decis că analiza situaţiei şi determinarea capacităţii necesare trebuie făcută cu maximum de atenţie şi competenţă, mai multă poate ca niciodată. Relaţiile dintre direcţiunea companiei şi grupul de cercetări operaţionale erau foarte cordiale, şi de aceea nu a constituit deloc o surpriză faptul că analiştii de cercetare operaţională au fost invitaţi să studieze capacitatea şi cerinţele operaţionale ale noului sediu şi să furnizeze date şi

recomandări ce puteau fi folosite ca o bază de discuţie cu arhitecţii şi cu autorităţile oraşului.

Primul pas făcut de personalul de cercetare operaţională a fost de a observa operaţiile ce se executau în sediul provizoriu, pentru a putea identifica principalele probleme ale sediului şi de a culege date statistice referitoare la acestea. Aceste studii au început la 29 aprilie 1958. Ele au progresat aproximativ o săptămînă, pînă cînd o grevă a personalului din transporturile londoneze a determinat oprirea activităţii la sediu; toate operaţiile au fost transferate la aeroport. Studiul a reînceput în iunie atunci cînd greva a luat sfîrşit. O perturbaţie de o asemenea natură are efecte profunde asupra validităţii datelor, şi echipa de cercetători trebuie să-şi ia măsuri suplimentare de precauţie atunci cînd le analizează. Pe parcursul analizelor efectuate au fost identificate 11 sisteme de aşteptare. Trei dintre ele au fost asociate cu operaţiile de verificare, asemănătoare celor discutate în capitolele anterioare.

Am selectat trei din celelalte pentru a le discuta în cadrul acestui capitol. Ele sînt:

1. Problema parcării maşinilor pentru o perioadă scurtă.
2. Sistemul de împrăştiere a clienţilor.
3. Problema încărcării autobuzelor la plecare.

13.2. CAPACITATEA LOCULUI DE PARCARE

Conceptual, aceasta a fost cea mai simplă problemă; de aceea este util să o discutăm la început. O trecere în revistă a modului în care pasagerii sosesc la sediul companiei de la casele lor sau de la hoteluri arată că 11,3% vin cu maşina personală. Adică, ei sînt conduşi de către soţiile lor, de către prieteni sau în unele cazuri de şoferi. Mai mult, 5,5% din pasagerii ce sosesc la sediul companiei de la aeroport îl părăsesc cu maşina lor personală. O observare atentă arată că, aproape fără excepţie, şoferii maşinilor particulare, ce însoţesc sau aşteaptă pasagerii la sediu, doresc să-şi parcheze maşina cît mai aproape de sediu, parcarea făcîndu-se pentru circa 20—30 minute în jurul momentului de plecare sau sosire a autobuzului.

Aceste necesităţi de parcare nu erau şi nu puteau fi satisfăcute de vechiul sediu. Spaţiul de parcare existent era foarte limitat de poziţia

clădirii. Sigur că nu era în intenţia companiei să nu creeze facilităţile necesare pasagerilor şi prietenilor lor ; din contră, dar nu era suficient spaţiu. S-a prevăzut totuşi ca noul sediu să prezinte mari posibilităţi de parcare pe termen scurt. Unele dintre aceste locuri de parcare ar putea fi foarte ieftine, în timp ce altele necesită fonduri mari, care ar putea fi valorificate mai bine.

Problema era de a stabili numărul maşinilor ce ar putea fi parcate lîngă noul sediu ; sau, pentru a privi altfel problema, să se calculeze proporţia celor care au nevoie de spaţiu de parcare şi nu găsesc loc (din care cauză vor căuta pe străzile învecinate sau vor veni la sediu cu taxiul).

Companiile aeriene există şi se dezvoltă în strînsă legătură de gradul de satisfacere a cerinţelor călătorilor (şi bineînţeles, gradul de securitate pe care îl garantează). Ar fi dificil de imaginat o altă întreprindere industrială în care se cheltuieşte atît de mult timp în legătură cu acest aspect al activităţii sale. Prin urmare, problema ce se punea lua forma următoare : cîte locuri de parcare sînt necesare, astfel încît la orele de vîrf ale activităţii, pentru luna iulie, în primul, al cincilea şi al zecilea an de la proiectarea sediului, numai o maşină din 20 sosite să nu poată intra imediat în locul de parcare. În discuţii s-a cerut la început ca nici o maşină sosită să nu aştepte eliberarea unui loc de parcare, ci să găsească repede un asemenea loc. Pe de-o parte, limitarea de spaţiu nu poate permite acest lucru, dar, pe de altă parte, dacă se alocă spaţiu pentru pasagerii care aşteaptă zborurile, atunci este destul de evident că trebuie alocat spaţiul corespunzător pentru parcare.

Problema, în limbajul teoriei aşteptării, nu era o problemă a timpului de aşteptare, ci una de *pierdere*. Iar măsura eficacităţii soluţiei nu era dată de vreun parametru al repartiţiei timpului de aşteptare, ci era tocmai probabilitatea respingerii, adică este probabilitatea ca o maşină sosită să fie nevoită să se întoarcă, deoarece locul de parcare este plin.

Această operaţie poate fi tratată ca o problemă de aşteptare. Staţiile de servire sînt locurile de parcare ; sosirile sînt maşinile căutînd un loc de parcare ; timpul de servire este tocmai timpul de ocupare a spaţiului de parcare ; nu este permisă aşteptarea. Se remarcă faptul că este un model extrem de uşor de mînuit, deoarece atîta timp cît sosirile sînt repartizate Poisson (aşa cum s-a constatat că sînt în acest caz) probabilitatea ca toate staţiile de servire să fie ocupate este independentă de forma exactă a repartiţiei timpului de

servire, atîta timp cît este de tipul E_k, şi este dependentă numai de timpul mediu de servire. Cititorul care doreşte demonstraţia celor afirmate este rugat să urmărească cap. 4, cazul mai multor staţii de servire. Formula este (v. cap. 4)

$$p_c = p_0 \frac{(c\rho)^c}{c!}, \tag{13.1}$$

unde p_0 este probabilitatea ca toate cele c staţii de servire să fie libere, iar ρ pentru acest caz cu mai multe staţii de servire este $\lambda/(c\,\mu)$. Formula pentru p_0 este

$$p_0 = \frac{1}{\sum_{r=0}^{c} \frac{(c\rho)^r}{r!}}. \tag{13.2}$$

Singura problemă ce urma să fie rezolvată era următoarea : timpul pentru care pasagerii doresc să-şi parcheze maşina este sau nu repartizat Erlang. Aceasta nu este doar o simplă chestiune de observaţie, deoarece repartiţia timpului de parcare poate fi controlată prin adoptarea unei politici corespunzătoare a tarifului de parcare. Modul general al politicii de tarifare pentru o operaţie ca cea a parcării pe o durată scurtă de timp este următorul :

pentru parcare mai puţin de T_0 min :	nici o taxă,
parcare între T_0 şi T_1 min :	M_1 şilingi/h,
parcare mai mult de T_2 min :	M_2 şilingi/h.

M_1 este o taxă care nu se ridică la un nivel foarte ridicat — să spunem 5 şilingi, funcţionînd ca un îndemn adresat şoferului pentru a părăsi locul. M_2 este o taxă suficient de ridicată, menită a penaliza şoferul egoist (care ignorează necesităţile şi drepturile celorlalţi).

A fost necesar să se realizeze acest sistem de taxe. Nu a fost o chestiune uşoară şi dealtfel nici nu am realizat-o extrem de bine. Principalul rezultat pe care l-am obţinut a fost acela că majoritatea călătorilor s-au obişnuit să utilizeze doar timpul rezervat lor, nu să-l depăşească. Am recomandat apoi ca taxa M_1 să fie ridicată ; presupunînd că ea este suficient de mare, putem continua calculele presupu-

nînd că timpul de servire (adică parcarea) este constant şi egal cu limita T_0. Repartiţia timpului constant de servire este un caz particular al repartiţiilor Erlang R_k, anume cazul în care k este infinit. Pot fi aplicate formulele (13.1) şi (13.2) şi se obţine numărul de locuri de parcare. Aceasta a fost o parte a proiectului ce nu a fost folosită.

Învăţătura rezultată din acest scurt exemplu este aceea că pentru problemele de proiectare şi de planificare în care trebuie aplicată teoria aşteptării este necesar să cunoaştem ce repartiţie a timpului de servire va fi. Ar fi poate mai corect — din punctul de vedere al preciziei — să studiem diferite variante alternative decît a elabora un model al sistemului de aşteptare.

O analiză exhaustivă a problemei parcării a fost realizată de Hurst în cadrul portului New York.

13.3. ÎMPRĂŞTIEREA PASAGERILOR CE PĂRĂSESC SEDIUL

Vom prezenta în cele ce urmează situaţia existentă în 1958. Pasagerii ce soseau la aeroportul Heathrow cu avioanele companiei BEA erau, în cazul cînd doreau acest lucru, duşi la Londra cu autobuzul companiei. La sediul companiei, pasagerii descindeau din autobuz, îşi aşteptau bagajele şi apoi fiecare se îndrepta spre oraş, utilizînd diferite mijloace : taxi, metrou, autobuz, pe jos etc.

Majoritatea pasagerilor preferau să părăsească sediul cu taxiul, iar în perioada pentru care s-a realizat studiul, s-a putut remarca că la anumite ore pasagerii soseau de la aeroport mai repede decît puteau fi preluaţi de către taxiuri. Din acest motiv ieşirile sediului companiei erau foarte congestionate.

Problema era de a determina, pentru viitorul sediu al companiei, nivelul de preluare al pasagerilor de către taxiuri, în scopul eliminării inconvenientului amintit. Problema avea mai multe părţi, pe care le vom prezenta sub forma unor întrebări :

1. Cîţi pasageri vor dori să părăsească sediul în orele de vîrf în anii ce urmează ?
2. Ce procent din aceştia vor dori să părăsească sediul cu taxiul ?
3. Cîţi pasageri vor fi aduşi de la aeroport cu autobuzul, în orele de vîrf ?

4. La ce intervale de timp vor sosi autobuzele la sediu, în orele de vîrf, în următorii ani?
5. Cîţi pasageri sînt transportaţi de un taxi?
6. Cît timp necesită încărcarea bagajelor şi încărcarea pasagerilor într-un taxi?

La prima din aceste întrebări s-a răspuns cu ajutorul prognozelor de trafic, realizate de sectorul de planificare al companiei. Nu este cazul să prezentăm aici aceste prognoze.

La a doua problemă răspunsul a fost dat astfel: doi reporteri au călătorit cu autobuzele între aeroport şi sediu timp de mai multe zile, chestionîndu-i pe pasageri asupra modului în care ei intenţionează să părăsească sediul cît şi asupra modului în care ar dori ei să o facă dacă sediul ar fi legat direct, prin intermediul unei scări rulante, de staţia de metro. Răspunsurile în cele două cazuri nu diferă foarte mult, şi datorită faptului că s-a hotărît să nu se construiască o legătură directă spre staţia de metro, au fost luate în considerare doar următoarele:

Procentul pasagerilor care susţin că vor părăsi sediul:

cu taxiul	74,0;
pe jos pînă la staţia de metro şi apoi cu metroul	16,5;
cu maşina personală:	5,5;
pe jos:	2,5;
cu autobuzul:	1,5.

Nu a surprins popularitatea taxiurilor. Ceea ce a surprins a fost gradul mic de variaţie a procentelor celor ce au preferat taxiurile în diversele curse investigate. De fapt, pentru 95% dintre autobuze, acest procentaj este cuprins între 71 şi 79 %. Rezultatele analizei făcute au fost confirmate, prin observări directe a comportamentului călătorilor în cadrul fostului sediu.

Nu era dificil de stabilit nici numărul pasagerilor ce sosesc cu autobuzul în orele de vîrf prin observări directe; din fericire, înainte ca această muncă să capete amploare, s-a ajuns la concluzia că ea nu era necesară. Numărul pasagerilor ce iau loc într-un autobuz variază foarte mult în cadrul unei zile, dar este destul de evident că el nu variază prea mult în timpul orelor de vîrf. Pe parcursul unor asemenea perioade, dispecerii auto de la aeroport nu lăsau autobuzele să plece avînd locuri libere; de cele mai multe ori ei aşteptau ca un autobuz

să se umple şi abia după aceea îl lăsau să pornească spre sediu. În perioadele în care cursele soseau foarte frecvent, iar numărul pasagerilor era extrem de ridicat, autobuzele se încărcau rapid şi această politică nu reprezenta o întîrziere sau un inconvenient pentru pasageri. De aceea atît atunci, cît şi pentru viitor, s-a considerat că numărul pasagerilor ce iau loc într-un autobuz, în orele de vîrf, este constant. A fost foarte uşor de determinat, prin observări directe, numărul pasagerilor ce iau loc într-un taxi : media era 1,35, iar repartiţia foarte concentrată pe valorile 1 şi 2.

Avînd în vedere că numărul pasagerilor ce solicită taxiurile este relativ constant, constanţa încărcării autobuzelor în orele de vîrf şi gradul scăzut de variaţie a numărului de pasageri ce iau loc într-un taxi, nu greşim prea mult aproximînd repartiţia φ_r a numărului de taxiuri corespunzătoare unui autobuz plin, în orele de vîrf. O calculăm aproximînd prima şi a treia componentă cu ajutorul repartiţiei binomiale; dar în final ne-am hotărît s-o considerăm constantă, avînd o valoare suficient de mare, cam de nivelul 80%-cuantilei. În consecinţă am considerat că fiecare autobuz încărcat generează o cerere de N taxiuri.

În timpul orelor de vîrf plecarea autobuzelor de la aeroport are loc, în conformitate cu politica de încărcare şi cu densitatea sosirii avioanelor, la intervale aproximativ regulate. Sosirile la sediu erau, însă, destul de neregulate. Figura 13.1 prezintă repartiţia duratelor dintre sosirile autobuzelor, în timpul orelor de vîrf, doar pe parcursul studiilor iniţiale. Fără îndoială că aceasta nu este constantă. Era limpede că sursa neregularităţilor era sistemul de şosele dintre Heathrow şi sediul din Kensington. Există o situaţie specială la Hammersmith, care acţiona asupra autobuzelor ca un concentrator, adunîndu-le laolaltă. Au fost consultate autorităţi competente, cum ar fi poliţia metropolitană şi laboratorul de cercetări rutiere. Rezultatele acestor consultări au fost acelea că programul de îmbunătăţiri rutiere prevăzute pentru cartierele vestice ale Londrei vor elimina aceste dificultăţi în cîţiva ani.

Fiind însărcinaţi cu studiul de perspectivă, am considerat că în anii următori repartiţia timpilor dintre sosirile autobuzelor în orele de vîrf va fi constantă. Era evident că intervalele de timp dintre sosiri nu vor fi constante, dar şi mai evident era faptul că ele vor fi mai mult constante decît aleatoare. Trebuie să recunosc că am greşit, dar extrem de puţin.

Este posibil ca unii cititori să fie în acest moment foarte nedumeriți. Abordarea clasică a problemei, așa cum este indicată în cărțile didactice de cercetare operațională, trebuie să fie cu totul alta. Unde sînt oare simulările, fotografiile aeriene ale congestiilor de trafic pe șoseaua Cromwell, analizele amănunțite ale „concentratorului" de la Hammersmith prin analogie cu hidrodinamica? Trebuie să mărturisesc că nu le-am considerat necesare și nu ne-am permis să cheltuim timpul și banii ca să le facem.

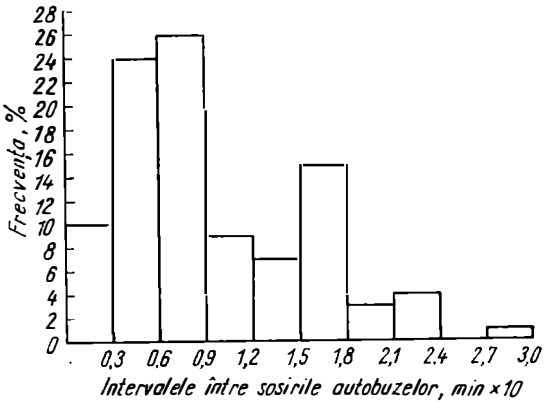


Fig. 13.1. Variația intervalelor dintre sosirile consecutive ale autobuzelor, cu media de 9 min, numărul de observație fiind egal cu 68.

În fine, repartiția timpului necesar unui taxi pentru încărcarea bagajelor, urcarea pasagerilor și părăsirea peronului a fost obținută prin cronometrare directă. Ceea ce rezultă poate fi considerat ca timpul pentru care un taxi exclude pe altul de la folosirea peronului, deci nu este un timp de servire normal, ci un timp de *blocare*. Repartiția empirică obținută este prezentată în fig. 13.2. Odată cu aceasta, echipa de cercetători a fost capabilă să construiască modelul sistemului de împrăștiere a pasagerilor cu ajutorul taxiurilor. Caracteristicile principale ale modelului erau:

1. Intrările în sistem: la fiecare x minute sosește un autobuz la sediu și pasagerii săi au nevoie de taxiuri.

2. Un autobuz plin de pasageri, în orele de vîrf, generează o cerere de N taxiuri, unde N este un parametru luînd valori întregi,

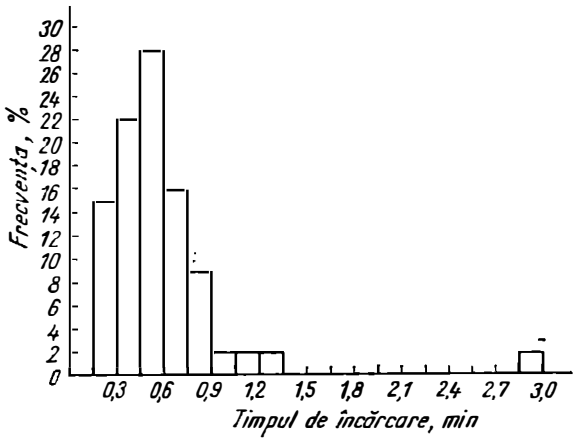


Fig. 13.2. Timpul necesar pentru urcarea pasagerilor în taxiuri, cu timpul mediu de încărcare de 0,5 min, numărul de observaţii fiind egal cu 54.

ce are posibilitatea de a varia anual, dar este constant pentru o perioadă de vîrf dintr-un anumit an.

3. Există un spaţiu adecvat pentru încărcarea simultană a c taxiuri.

4. Timpul de *blocare* (de încărcare) al unui taxi este t, avînd densitatea de repartiţie $b(t)$.

5. Ordinea în care pasagerii sau grupurile de pasageri iau taxiurile este nesemnificativă.

Problema era de a stabili repartiţia timpului total necesar celor N taxiuri ca să preia toţi pasagerii unui autobuz.

Obiectivul era de a determina un c suficient de mare pentru a asigura cu o probabilitate mare ca toţi pasagerii dintr-un autobuz să plece din sediu înaintea sosirii următorului autobuz.

Este suficient să tratăm operţiunea ca desfăşurîndu-se în două faze. În prima fază, numărul taxiurilor ce urmează să fie încărcate, pe care-l vom nota cu $N - r$, nu este mai mic decît c, numărul persoanelor. În faza a doua, acest număr este mai mic decît c.

Presupunem că densitatea de repartiţie a duratei t_1, care este timpul cuprins între sosirea pasagerilor şi momentul în care rămîn $c - 1$ taxiuri de încărcat, este $f(t_1)$; fie $g(t_2)$ densitatea de repartiţie a duratei, necesare încărcării acestor $c - 1$ taxiuri. În acest caz,

timpul total t de la sosirea pasagerilor şi pînă la terminarea încărcării ultimului taxi, are densitatea de repartiţie dată de

$$h(t) = \int_0^t f(t_1)\, g(t - t_1)\, \mathrm{d}t_1.$$

Această relaţie este mult simplificată utilizînd cunoscuta prioritate a transformatelor Laplace :

$$h^*(z) = f^*(z) \cdot g^*(z). \tag{13.3}$$

Trebuie menţionat că $h(t)$ va fi dependentă de N, atîta timp cît nu s-a făcut vreo ipoteză asupra lui N.

Următorul pas este de a determina funcţiile $f(t_1)$ şi $g(t_2)$. Le vom determina utilizînd două modele simple. Vom începe cu un model pentru prima fază. Acest model presupune că probabilitatea eliberării unui peron prin încărcarea unui taxi într-un interval de timp $\mathrm{d}t$ este $\mu \mathrm{d}t$, astfel încît atîta timp cît numărul taxiurilor ce rămîn să fie încărcate este mai mare decît c, intensitatea medie de încărcare a taxiurilor este $c\mu$. Adică, repartiţia timpilor de blocare (de încărcare) este exponenţială, avînd forma

$$\mu \mathrm{e}^{-\mu t}.$$

Deşi acest lucru nu este întru totul adevărat, aşa cum reiese din fig. 13.2, totuşi este o primă aproximaţie suficient de corectă. Timpii reali de încărcare nu sînt atît de aleatori aşa cum reiese din această presupunere, astfel încît utilizînd-o vom avea tendinţa de a supraestima numărul peroanelor necesare. Pasagerii vor părăsi sediul mult mai rapid decît rezultă din calcule.

Fie $p(N - r, t)$ probabilitatea ca la un moment t să rămînă $N - r$ taxiuri ce urmează a fi încărcate cu pasageri. Atunci pentru prima fază avem

$$\frac{\mathrm{d}}{\mathrm{d}t} p(N, t) = - c\mu p(p(N, t),$$

$$\frac{\mathrm{d}}{\mathrm{d}t} p(N - r, t) = - c\mu p(N - r, t) + c\mu p(N - r + 1, t) \quad (0 < r \leqslant c),$$

$$\frac{\mathrm{d}}{\mathrm{d}t} p(c - 1, t) = c\mu p(c, t).$$

Soluţia acestor ecuaţii este

$$f(t_1) = c\mu e^{-c\mu t_1} \frac{(c\mu t_1)^{N-c}}{(N-c)!}. \tag{13.4}$$

Transformata Laplace a acestei funcţii este

$$f^*(z) = \left(\frac{c\mu}{c\mu + z}\right)^{N-c+1}. \tag{13.5}$$

Modelul utilizat pentru a doua fază presupune de asemenea încărcări aleatoare. Cititorul va identifica cu siguranţă ecuaţiile modelului, care sînt:

$$\frac{\mathrm{d}}{\mathrm{d}t} p(c-1, t) = -(c-1)\mu p(c-1, t),$$

$$\frac{\mathrm{d}}{\mathrm{d}t} p(r, t) = -r\mu p(r, t) + (r+1)\mu p(r+1, t) \quad (1 \leqslant r < c-1),$$

$$\frac{\mathrm{d}}{\mathrm{d}t} p(0, t) = \mu p(1, t).$$

Soluţia se găseşte uşor:

$$g(t_2) = \binom{c-1}{c-2} e^{-\mu t_2} (e^{-\mu t_2} - 1)^{c-2} \mu. \tag{13.6}$$

Transformata Laplace a acestei expresii este

$$g^*(z) = \frac{1}{(1 + z/\mu)} \cdot \frac{\Gamma(1 - z/\mu)\,\Gamma(c-2)}{\Gamma(c - z/\mu - 1)}. \tag{13.7}$$

Utilizînd cunoscuta identitate $\Gamma(z) = (z-1)\,\Gamma(z-1)$, aceasta poate fi scrisă ca

$$\frac{1}{\left(1 + \frac{z}{\mu}\right)\left(1 + \frac{z}{2\mu}\right)\left(1 + \frac{z}{3\mu}\right) \ldots \left(1 + \frac{z}{(c-1)\mu}\right)}.$$

În consecință, transformata Laplace a repartiției timpului total de încărcare a N taxiuri, a cărei densitate este $h(t)$, este dată de

$$h^*(z) = \left\{\prod_{r=1}^{c-1}\left(1+\frac{z}{r\mu}\right)\right\}^{-1}\cdot\left(1+\frac{z}{c\mu}\right)^{-N+c-1}. \tag{13.8}$$

Ea poate fi scrisă și sub forma :

$$\left\{1-\frac{z}{\mu}+\frac{z^2}{\mu^2}+\ldots\right\}\left\{1-\frac{z}{2\mu}+\frac{z^2}{4\mu^2}+\ldots\right\}\ldots$$

$$\ldots\left\{1-\frac{z}{(c-1)\mu}+\frac{z^2}{(c-1)^2\mu^2}+\ldots\right\}\left\{1-(N-c+1)\frac{z}{c\mu}+\ldots\right\}.$$

Dar transformata Laplace este o funcție generatoare, astfel încît coeficienții lui $\frac{z^r}{r!}$ luați cu semne corespunzătoare sînt momentele m_r' în vecinătatea originii ale repartiției $h(t)$. Acestea pot fi ușor modificate cu ajutorul formulelor

$$m_2 = m_2' - m_1'^2; \qquad m_3 = m_3' - 3m_1'm_2 - m_1'^3, \ldots,$$

ce le convertesc în momente în vecinătatea valorii medii. Primele patru asemenea momente sînt :

$$m_1 = \left[\sum_{r=1}^{c-1}\frac{1}{r}+\frac{N-c+1}{c}\right]\frac{1}{\mu}, \tag{13.9}$$

$$m_2 = \left[\sum_{r=1}^{c-1}\frac{1}{r^2}+\frac{N-c+1}{c^2}\right]\frac{1}{\mu^2}, \tag{13.10}$$

$$m_3 = \left[\sum_{r=1}^{c-1}\frac{1}{r^3}+\frac{N-c+1}{c^3}\right]\frac{1}{\mu^3}, \tag{13.11}$$

$$m_4 = \left[\sum_{r=1}^{c-1}\frac{1}{r^3}+\frac{N-c+1}{c^4}\right]\frac{6}{\mu^4}+\left[\sum_{r=1}^{c-1}\frac{1}{r^2}+\frac{N-c+1}{c^2}\right]\frac{3}{\mu^4}. \tag{13.12}$$

Cititorul va remarca din aceste momente că coeficienții Pearson de asimetrie și exces sînt respectiv :

$$\sqrt{\frac{m_3^2}{m_2^3}} = o\left(\frac{1}{\sqrt{N}}\right), \quad \frac{m_4}{m_2^2} - 3 = o\left(\frac{1}{N}\right).$$

Așadar, pentru valori suficient de mari ale lui N, repartiția este aproximativ normală. Este un lucru de mare importanță practică, deoarece inversarea transformatei Laplace $h^*(z)$ este o sarcină foarte dificilă.

Numărul de peroane necesare pentru încărcarea pasagerilor în taxiuri, peroane care în acest caz sînt tocmai stațiile de servire, a fost calculat pentru fiecare din anii care interesează analiza, folosind această aproximare normală. Altfel spus, a considera că cu o probabilitate de 95% pasagerii unui autobuz ce au nevoie de N taxiuri se vor împrăștia înaintea sosirii următorului autobuz revine la a calcula pe c astfel încît

$$m_1(c) + 1{,}96\sqrt{m_2(c)} \leqslant x. \tag{13.13}$$

Calculele făcute nu vor fi prezentate aici, deoarece o prezentare detaliată va cere explicații privind prognoza traficului. În orice caz, aceste chestiuni nu prezintă interes.

Rezultatele acestor calcule au fost utilizate în proiectarea sediului. Înainte de a părăsi acest proiect al sediului companiei, ar fi util să facem o sumară trecere în revistă a aspectelor legate de modelul ce ar fi rezultat, dacă nu am fi făcut presupunerea simplificatoare că fiecare autobuz generează un număr constant de taxiuri în timpul orelor de vîrf. Este evident că în realitate acest număr, pe care-l notăm cu n, nu este constant. El are o repartiție nesimetrică care nu poate fi identificată cu precizie. Am fi putut repartiza probabilități pentru diferite valori întregi (concentrate cu o mare frecvență în jurul mediei) definind astfel o funcție de repartiție, să spunem φ_n. În acest caz transformata Laplace a repartiției timpului de încărcare a pasagerilor în taxiuri va fi următoarea :

$$\sum_{n=0}^{M} \varphi_n h_n^*(z),$$

unde $h_n^*(z)$ este identică cu $h^*(z)$, în care N este înlocuit cu n, iar M este numărul pasagerilor transportaţi de către un autobuz. O astfel de analiză ar fi însemnat o muncă uriaşă, fără îndoială lăudabilă, dar practic inutilă. Calculele menţionate nu sînt imposibil de realizat, dar necesită o muncă foarte laborioasă.

13.4. ÎNCĂRCAREA AUTOBUZELOR SPRE AEROPORT

Pasagerii care au fost verificaţi pentru zbor la sediul companiei aeriene trebuie transportaţi la aeroport pentru a se alătura celor care au fost verificaţi direct la aeroport. La aeroport, cele două grupuri fuzionează. Există mai multe modalităţi de a organiza această operaţie. Din raţiuni pe care nu este necesar să le discutăm acum, s-a decis să fie menţinută metoda utilizată la sediul existent. Adică fiecărui avion care decolează de la aeroport îi va corespunde un autobuz care va părăsi sediul la o oră fixată, înaintea orei de decolare. Durata deplasării (pe atunci de 75 min) depinde de condiţiile rutiere dintre Kensington şi Heathrow; se estima că această durată se va micşora. Durata exactă nu ne interesează, doar că ea determină planificarea exactă a autobuzelor de la sediul companiei, ca o cacaracteristică a metodei utilizate.

Autobuzele nu pot pleca spre aeroport mai devreme de ora fixată (există cazuri cînd apar pasageri în ultimul moment), iar uneori ele pleacă mai tîrziu (datorită întîrzierilor în verificarea pasagerilor sau în sortarea bagajelor). Figura 13.3 arată repartiţia întîrzierilor plecărilor de la sediul existent. Nu ne aşteptam la o schimbare a acestui fenomen în viitor, oricum însă unele modificări aduse vor îmbunătăţi sistemul.

Timpul necesar urcării pasagerilor într-un autobuz era dependent într-un mod oarecare de numărul acestora. Şi datorită măririi capacităţii avioanelor pe liniile BEA, ne aşteptam ca numărul pasagerilor ce trebuiau transportaţi cu autobuzul să sporească; trebuiau găsite anumite modalităţi, de exemplu trebuiau mărite dimensiunile autobuzelor de transport, dar trebuia să găsim şi o modalitate de a estima durata de încărcare a unui autobuz. Cercetătorii, membri ai echipei de studiu, au înregistrat timpul necesar încărcării pasagerilor din momentul anunţării îmbarcării şi pînă în momentul în care ultimul pasager s-a urcat în autobuz şi şi-a ocupat locul. Rezultatele sînt prezentate în fig. 13.4, care arată faptul că timpul de încărcare are o tendinţă de creştere odată cu creşterea numărului pasagerilor, dar

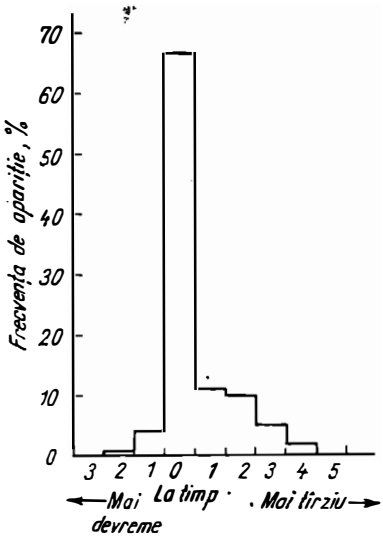


Fig. 13.3. Punctualitatea plecării autobuzelor, cu abaterea medie de 0,7 min, numărul de observații fiind egal cu 83.

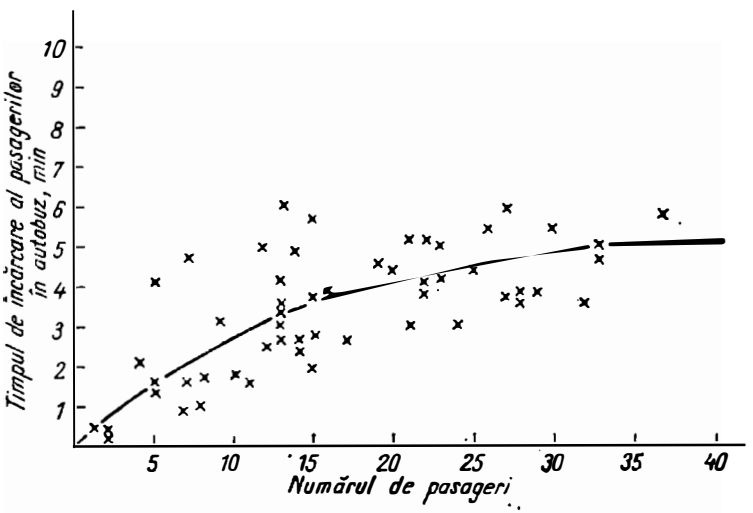


Fig. 13.4. Corelația dintre timpul de urcare al pasagerilor în autobuz și numărul de pasageri.

nu direct proporţională. Timpul mediu de încărcare a fost calculat aşa cum reiese din figură, iar abaterea standard a fost de asemenea estimată.

Analizele efectuate au arătat că repartiţia pasagerilor între sediu şi aeroport (adică cîţi vin la sediu la verificare şi cîţi direct la aeroport) a fost pe parcursul perioadelor de vîrf din ultimii doi ani o repartiţie binomială. Parametrul binomial α, reprezentînd probabilitatea alegerii sediului, variază sezonier şi zilnic, dar în timpul orelor de vîrf are o valoare relativ constantă. Prin urmare, numărul pasagerilor care solicită autobuzul la sediu, considerînd că sînt N pasageri înregistraţi pentru zborul respectiv, este dat de

$$P(r) = \binom{N}{r} \alpha^r (1 - \alpha)^{N-r}.$$

(Putem să facem abstracţie de pasagerii care nu au bilete rezervate dinainte, ducîndu-se direct la aeroport. Numărul acestora este foarte scăzut şi de regulă ei merg direct la aeroport.) Repartiţia numărului de pasageri pe zbor era cunoscută, şi în plus se cunoşteau şi previziunile pentru viitor. Aceasta a fost utilizată împreună cu repartiţia binomială prezentată mai sus şi cu repartiţia timpului de încărcare a pasagerilor (rezultată din datele din fig. 13.4), în scopul obţinerii unei repartiţii a timpului de încărcare a unui autobuz pe zbor. Au fost utilizate metodele de simulare Monte Carlo pentru a combina aceste trei repartiţii. Datorită faptului că în acea perioadă echipa de analişti nu putea folosi nici un calculator electronic, totul a fost realizat manual, utilizînd tehnici destul de învechite.

Nu ne-a interesat să observăm direct timpul necesar autobuzelor de a manevra spre şi de la locul de încărcare, deoarece ştiam că în noul sediu lucrurile vor fi cu totul diferite. Consultîndu-ne cu şeful proiectului noului sediu şi cu doi experţi, a fost posibil să obţinem estimări pentru acest timp de manevrare.

Timpul minim pentru care un autobuz blochează utilizarea locului de îmbarcare, adică împiedică pe următorul să-l folosească, a fost calculat adunînd-se următoarele componente :

1) timpul necesar autobuzului pentru manevrare spre locul de îmbarcare ;

2) timpul de îmbarcare a pasagerilor ;

3) întîrzieri ale plecării autobuzelor datorită pasagerilor ;

4) timpul necesar autobuzului pentru a ieşi de la locul de îmbarcare.

Cititorul poate să considere desigur că repartiţia formată de aceste componente este o repartiţie a timpului de servire. Sistemele de aşteptare, din categoria cărora face parte acesta, au ca staţii de servire locurile de încărcare. Dar cine vor fi clienţii? Clienţii vor fi autobuzele ce ajung la locurile de îmbarcare pregătite pentru a lua pasageri. Momentele în care autobuzele sînt gata de a încărca pasagerii trebuie planificate înaintea momentelor planificate pentru plecarea autobuzelor; ele trebuie să fie corelate cu momentele în care li se anunţă pasagerilor plecarea autobuzului.

Pe parcursul orelor de vîrf, aceste momente au fost dispuse regulat, atît la vechiul sediu, cît şi la cel nou. Intervalul constant dintre plecările autobuzelor, care poate fi prognozat cu cîţiva ani înainte, asigură acest lucru. De aceea intrările în sistemul de aşteptare sînt caracterizate prin sosiri constante, regulate. Se părea că modelul ce putea fi aplicat aici în mod corespunzător era modelul de aşteptare $D/E_k/c$: (∞/FIFO). Pentru echipa de cercetători era însă limpede că a trata problema într-o astfel de manieră nu era deloc util, deoarece încă nu erau deduse formulele pentru repartiţia timpului total de aşteptare pentru un astfel de model.

Ar fi fost posibil ca echipa de cercetători să fi obţinut rezultate satisfăcătoare prin analizarea teoretică a modelului $D/E_k/c$: (∞/FIFO), dar problema practică necesita o soluţionare urgentă. Era necesar a găsi cît mai repede un răspuns la întrebarea: „Cîte peroane de încărcare a autobuzelor vor fi necesare în 1963, 1965, 1970?".

Un răspuns exact, bazat pe calcule matematice, nu putea fi obţinut în numai cîteva zile; de aceea soluţia era de a da un răspuns foarte aproximativ. Din această cauză s-a renunaţat la modelul matematic al sistemului, acesta fiind înlocuit de un model bazat pe simulare.

Simularea reprezintă o metodă utilă pentru rezolvarea problemelor ce nu pot fi tratate analitic, cu ajutorul experimentelor aleatoare puse în concordanţă cu o structură logică de decizie. În cap. 6 am analizat metodele de simulare iar cititorul care nu cunoaşte principiile generale ar trebui să citească mai întîi primele paragrafe ale capitolului 6.

Odată luată decizia de simulare a sistemului, a devenit necesar a se analiza eventualele modificări ce vor fi aduse, pe parcursul următorilor ani. Una din cele mai frecvente greşeli ce se fac în simulare este

aceea de a nu lăsa suficientă flexibilitate structurii. Am studiat apoi un număr de metode operaţionale, ce puteau da rezultate bune. Acestea au fost testate, comparîndu-le cu modul de lucru existent. Pentru a reduce numărul pasagerilor ce necesită un fotoliu, fiind în aşteptare în cadrul sediului, şi pentru a micşora întîrzierile plecării autobuzelor, s-a propus de către unul dintre supraveghetori ca fiecare autobuz corespunzător fiecărei curse aeriene să fie parcat cu 15 minute înainte de ora la care este programată plecarea, astfel încît pasagerii să poată să se îmbarce din timp, pentru a evita aglomeraţia din ultimele minute. Această schemă a fost testată timp de cîteva luni la vechiul sediu şi se pare că a dat rezultatele aşteptate. Au apărut însă două mari dezavantaje ale acestei scheme, ambele cît se poate de evidente, dar pierdute din vedere în entuziasmul general ce a însoţit introducerea schemei şi testarea ei :

1) era necesar un număr mai mare de autobuze ;

2) în orele de vîrf, erau necesare mai multe peroane de îmbarcare.

Întrebările la care trebuia găsit un răspuns în ambele cazuri era : „Cîte anume ?"

În acest scop s-a simulat două proceduri operaţionale, şi s-au comparat rezultatele. Simularea a fost realizată manual, utilizînd hîrtie, creioane şi tabele de numere aleatoare, după cum urmează :

1. Au fost create mai multe orare ale plecărilor autobuzelor pentru perioade de vîrf de trei ore. Acestea presupuneau (fapt rezultat din observaţii) că în prima oră intervalele dintre plecări au tendinţa de descreştere, în timpul celei de-a doua ore de vîrf ele sînt constante, iar în ultima au din nou tendinţa de a creşte.

2. A fost selectat un orar. Începînd cu plecarea cea mai timpurie, pentru fiecare plecare a unui autobuz, se extrage un număr aleator dintr-un tabel cu numere aleatoare şi prin alăturarea sa la repartiţia „timpului de servire", se generează astfel în mod aleator un timp de servire.

3. Se ţinea o evidenţă privind numărul de autobuze prezente la peroane în fiecare moment. Numărul maxim înregistrat arăta numărul peroanelor necesare, dacă autobuzele nu trebuiau să aştepte eliberarea unui peron.

4. Această procedură a fost repetată de mai multe ori, utilizînd diferite orare de plecare, generîndu-se diferite secvenţe de numere aleatoare.

Pentru fiecare variantă operaţională a fost obţinută astfel o repartiţie a peroanelor necesare în orele de vîrf. Numărul peroanelor

necesare a fost considerat a fi numărul necesar pentru a evita întîrzierile la 19 din 20 autobuze. Cititorul ar putea fi interesat să afle anumite date concrete. De exemplu, într-o oră de vîrf, în care autobuzele sînt planificate să plece din trei în trei minute, în care timpul mediu de blocare este de nouă minute iar încărcarea medie a unui autobuz este de 22 de pasageri, erau necesare patru peroane, dacă se adopta schema veche. Dar în noua schemă propusă, în care timpul mediu de blocare era de 16,5 minute, era nevoie de opt peroane, în aceleaşi condiţii. Acest caz prezentat este un exemplu al unei foarte importante probleme în planificarea viitoarelor operaţii, care poate apărea ca fiind complicată pentru a fi analizată matematic, însă pe care un om practic o rezolvă destul de uşor.

14

PROBLEMELE DE AȘTEPTARE ÎNTR-UN SISTEM DE CONTROL

Cititorul va fi iertat dacă cele discutate pînă acum, atît din punct de vedere teoretic cît și cel practic, l-au condus la ideea că sistemele de așteptare sînt toate doar o chestiune de factori ori de relații între factori, totul fiind măsurabil în unități de timp — secunde, minute, ore sau săptămîni. Timpii dintre sosiri, timpii de servire, timpii de așteptare și timpii totali de așteptare sînt toți de acest tip. Majoritatea problemeleor ce apar în studiul sistemelor de așteptare sînt legate de astfel de timpi ; există însă și excepții, iar acest capitol prezintă una dintre ele.

Să considerăm următoarea situație. D-na Evans se duce la magazin pentru a-și cumpăra o cutie cu șase perechi de ciorapi de nylon, preambalați, ce se vinde cu un preț acceptabil. După o așteptare de circa cinci minute, ea ajunge să fie servită. „Vă rog mărimea 10, nuanță deschisă", îi spune vînzătoarei. Aceasta este însă obosită și grăbită (căci mai sînt mulți clienți ce așteaptă). O va servi, fără îndoială, cu promptitudine pe d-na Evans, care va părăsi magazinul cu cutia de ciorapi. Ajungînd acasă și deschizînd cutia, ea constată că ciorapii au mărimea 12, nuanță închisă. Oare timpul ei de așteptare, sau timpul ei total de așteptare, este pe măsura calității serviciului prestat de către vînzătoare ?

În cadrul acestui capitol ne vom ocupa de o problemă în care serviciul furnizat de o stație de servire poate fi un serviciu eronat, defectuos, în care probabilitatea unui serviciu defectuos este dependentă de comportamentul sistemului de așteptare, astfel încît cri-

teriul de determinare a capacităţii staţiei de servire este în primul rînd de natură calitativă şi abia în al doilea rînd şi indirect un timp de aşteptare.

14.1. PREZENTARE A PROBLEMEI

Avioanele sînt costisitoare şi nu sînt rentabile decît în măsura în care ele realizează cursele cu capacitate completă. Atunci cînd cererea de transport este extrem de ridicată, timpul pe care avionul îl petrece pe sol trebuie să fie minim. Atunci cînd un avion aterizează, la sfîrşitul unei curse, el trebuie trimis în altă cursă în cel mai scurt timp posibil. Este important ca dirijarea să se facă nu numai rapid ci şi cît mai convenabil, căci orice întîrziere a unei plecări poate determina scăderea substanţială a utilizării parcului de avioane.

Dirijarea avioanelor constă în realizarea succesivă a mai multor acţiuni, ordonate într-un anumit fel. Dacă o acţiune se termină mai tîrziu decît este prevăzut, următoarea va fi de asemenea întîrziată.

Anumite companii aeriene au ajuns la concluzia că operaţia de retrimitere a avioanelor în cursă, cel puţin la marile aeroporturi, trebuie să fie foarte bine dirijată, astfel încît cauzele reale şi potenţiale ale întîrzierilor să fie detectate şi eliminate rapid. Soluţiile alese diferă ca natură cît şi ca eficacitate şi sînt în mare măsură determinate de diferite moduri ale organizării conducerii aeroporturilor.

Una dintre cele mai raţionale soluţii a fost adoptată de compania BEA. Aceasta a constat în favorizarea înfiinţării, la fiecare aeroport mai important, a unui organ de control, al cărui personal să cunoască în orice moment starea avioanelor, în timpul care este necesar pregătirii fiecărui avion pentru decolare. Conducerea credea, justificat, că procedîndu-se în alt mod ar fi fost dificil de coordonat diversele activităţi, foarte variate dealtfel, şi de luat deciziile necesare impuse de înlocuirea unor curse. Astfel, în 1955 coordonarea şi dirijarea activităţilor de planificare a zborurilor pentru multe aeroporturi din Anglia a fost realizată în cadrul companiei BEA prin înfiinţarea departamentului de control al BEA.

În primăvara lui 1956, s-au constatat anumite neajunsuri la aeroportul din Londra, în timpul perioadelor de vîrf. Simptomele erau : întîrzieri excesive în prelucrarea informaţiilor, transmiterea

de informații eronate, culminînd cu neefectuarea de către întregul grup a unor operații esențiale. În consecință, conducerea companiei BEA a cerut departamentului de organizare să efectueze un studiu cu scopul de a crește eficiența departamentului de control de pe lîngă aeroportul din Londra, astfel încît să se revină la un nivel calitativ ridicat. Problema fiind de teoria așteptării, ea a fost preluată de un grup de cercetare operațională, în particular de către autorul prezentei lucrări.

14.2. STUDIUL PRELIMINAR

Prima sarcină a fost de a stabili ce anume trebuia să realizeze departamentul de control și ce realiza de fapt. Aceasta părea la prima vedere o sarcină considerabilă, dar care a putut fi dusă ușor la îndeplinire datorită faptului că cei interesați erau tocmai conducătorii departamentului. Pe scurt, iată sintetizate rezultatele acestei faze preliminare : inițial scopul departamentului de control era de a coordona activitățile de pregătire pentru zbor și de a asigura ca fiecare decolare să se facă la momentul planificat. Se presupunea că acest lucru se va realiza prin controlul continuu efectuat asupra avioanelor, în pregătirea lor pentru curse, și prin înregistrarea și controlul parțial al pasagerilor în incinta aeroportului. Responsabilitatea departamentului în ceea ce privește punctualitatea plecărilor cerea și îndeplinirea unei a doua funcțiuni : a depista cauzele tuturor întîrzierilor și raportarea lor conducerii.

Coordonarea activităților de pregătire pentru zbor a fost efectuată după cum urmează. Fiecare avion ce urma să decoleze era vizitat periodic de un observator, care nota activitățile terminate și estima, pe baza experienței sale, care activități încă neefectuate ar putea produce întîrzieri. Aceste notări și observații erau transmise centrului de control prin intermediul unui aparat de radio-emisie-recepție.

Informațiile erau recepționate la departamentul de control de către controlorul de zboruri care înregistra încheierea activităților pregătitoare, semnala întîrzierile potențiale celor interesați și dirija activitatea observatorului.

Atunci cînd apărea ca inevitabilă o întîrziere, controlorul de zboruri decidea ce modificări sînt necesare în pregătirea pentru zbor și informa toate secțiile în legătură cu situația ivită și cu durata probabilă a întîrzierii.

Celelalte secţii comunicau cu controlorul de zboruri printr-un sistem de intercomunicaţii, după cum se arată în fig. 14.1. Acea parte a muncii de control ce era legată de fluxul pasagerilor prin incinta aeroporturilor era realizată de un grup cunoscut sub numele de AGC.

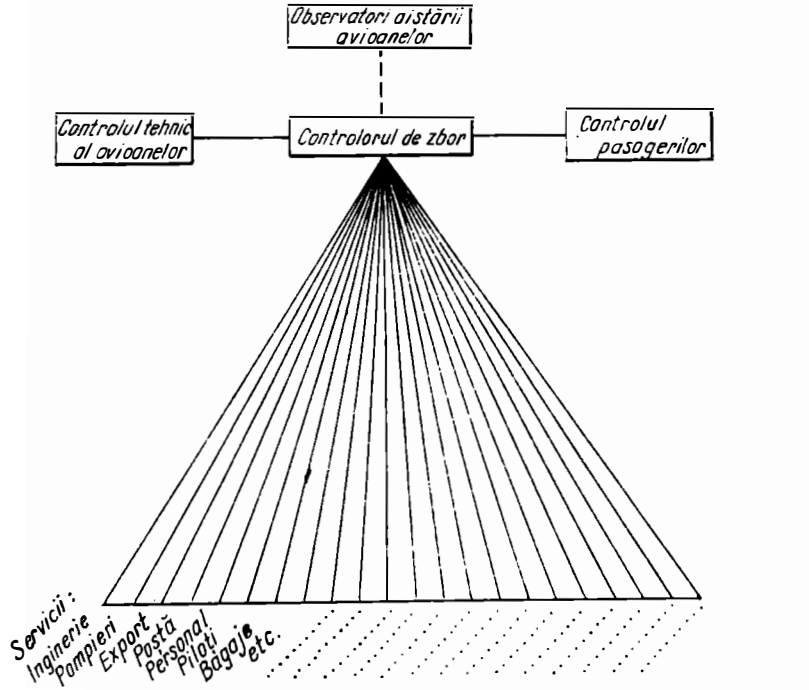


Fig. 14.1. Reţeaua de comunicaţie a controlorului de zbor.

Aceasta reprezenta grupul de legătură între controlorul de zboruri şi unităţile de servire a pasagerilor. Membrii acestui grup nu aveau autoritatea de a decide, îmbarcarea pasagerilor trebuind să se supună deciziilor luate de către controlorul de zboruri; ei trebuiau doar să controleze îmbarcarea corectă a pasagerilor şi să informeze unităţile de servire a pasagerilor în legătură cu orice întîrziere ivită.

Următoarea fază a studiului a constat în identificarea motivelor care provoacă întîrzierile. Aceasta au apărut cu claritate în timpul perioadelor de vîrf. Primul motiv a fost dificultatea de a se realiza legătura între diversele secţii şi controlorul de zbor pentru a se

schimba informaţii. Al doilea motiv era faptul că atît observatorii aparatelor de zbor cît şi membrii grupului AGC trebuiau să transmită datele şi problemele ce necesita luarea unei decizii controlorului de zbor, care, deşi experimentat, avea o serie de dificultăţi în luarea acestor decizii. În al treilea rînd, controlorul de zbor încurca şi uita unele informaţii, fapt dovedit de transmiterea unor instrucţiuni şi decizii eronate.

Aceste motive, analizate şi în cadrul departamentului de control, au arătat limpede că sursa deficienţelor era departamentul de control al plecărilor. Prin urmare, principalul cîmp de studiu era funcţionarea acestui departament de control.

14.3. STUDIUL PRINCIPAL: CONTROLUL PLECĂRILOR

Pentru a realiza studiul funcţionării departamentului la control a trebuit să selectăm un număr de ore de vîrf şi prin observare directă să determinăm conţinutul şi durata mesajelor transmise şi recepţionate către şi de la controlorul de zbor. Această fază a proiectului a stabilit că funcţia de control a zborurilor, efectuată de o singură persoană, nu era deloc simplă. Figura 14.2 arată modul de recepţionare a informaţiei prin sistemul de comunicaţie şi clasificarea acesteia după gradul de urgenţă şi după durata de folosire. Informaţia este

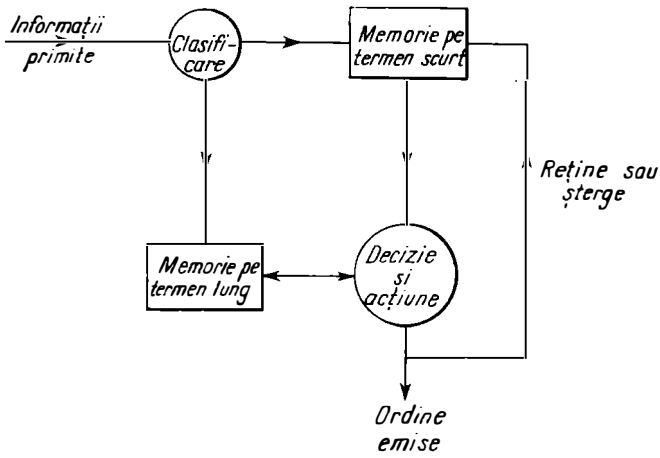


Fig. 14.2. Funcţia de control al zborurilor.

memorată prin scrierea ei pe un program al curselor aeriene la pupitrul controlorului de zbor, sau menţinută în memorie pînă cînd va fi solicitată pentru luarea unei decizii.

Memoria controlorului de zbor era asigurată într-un sistem de memorare cu acces direct pentru informaţiile cu termen scurt de păstrare şi printr-un alt sistem de memorare pentru informaţiile cu durată mai lungă de păstrare. O memorie pentru informaţiile cu termen scurt de păstrare a fost necesară din două motive : în primul rînd, ca o memorie tampon, pentru a reţine toate informaţiile mai rapid decît pot fi ele prelucrate şi, în al doilea rînd, ca un acumulator pentru a colecta informaţiile referitoare la o anumită problemă specifică, înaintea luării unei decizii. Scopul memoriei pentru informaţiile cu termen lung de păstrare era de a reţine care activităţi importante au fost terminate în cadrul operaţiilor de pregătire a avioanelor pentru decolare şi de a reţine informaţiile cu caracter statistic (de exemplu numărul de pasageri transportaţi).

Acest model al funcţionării controlorului zborurilor sugerează o analiză de aşteptare pentru a determina încărcarea controlorului de zbor în timpul perioadelor de vîrf. Observaţiile făcute în aceste perioade au arătat că controlorul era ocupat cu rezolvarea apelurilor într-o proporţie de 65 % din timpul său disponibil. Durata apelului, care în acest caz este mai curînd un timp de blocare decît un timp de servire (deoarece cuprinde cîteva secunde în plus de serviciu suplimentar după terminarea recepţionării unui mesaj) a fost găsită avînd o repartiţie aproximativ exponenţială, cu o medie de 0,23 min. S-a constatat că informaţiile erau tratate într-o ordine aleatoare. Modelul de aşteptare pentru controlorul de zboruri este — cel puţin în primă aproximaţie — obişnuitul $M/M/1$: (∞/SIRO), iar tabelele lui Wilkinson arătau că comunicările de informaţii sînt întîrziate în medie cu mai mult de 0,5 min, cel puţin 6 % dintre ele fiind întîrziate cu mai mult de 2,5 min. Aceste calcule au fost confirmate de verificările făcute. Deci controlorul de zboruri era suprasolicitat în orele de vîrf. În consecinţă, trebuie urmată una din următoarele două căi : 1) reducerea numărului de informaţii ce trebuie recepţionate şi prelucrate, sau 2) divizarea muncii între două sau mai multe persoane.

A fost realizat, cu destul de mult succes, un studiu privind reducerea numărului de informaţii. Acest studiu nu implică probleme de teoria aşteptării, aşa încît nu-i vom acorda atenţie. Rezultatul acestui studiu a fost reducerea numărului de informaţii primite într-o proporţie de 70 % comparativ cu nivelul anterior. Aceasta a micşorat

încărcarea controlului de zbor în orele de vîrf și ne-a permis să ne ocupăm de altă problemă. Întrucît pînă acum n-a fost vorba deloc despre informații prelucrate eronat sau uitate, am acordat atenție în continuare analizei erorilor făcute de către oameni supraîncărcați în orele de vîrf. O analiză a apariției erorilor în cadrul mesajelor, bazată pe informațiile stocate în memoria controlorului de zbor a arătat că în perioadele de vîrf frecvența erorilor era mai ridicată. Cu alte cuvinte, probabilitatea ca o informație să fie distorsionată sau pierdută crește odată cu creșterea încărcării controlorului de zboruri. Nu se cunoștea o formă exactă a acestei relații și ar fi foarte dificil să o determinăm, dar observațiile făcute sugerează că ar putea fi reprezentată aproximativ astfel :

$$f(n) = 0{,}005\ n, \tag{14.1}$$

$f(n)$ fiind probabilitatea ca o informație să fie pierdută sau distorsionată, iar n este numărul de informații primite pe minut și memorate pe termen scurt.

Cum un număr de n mesaje memorate pe termen scurt era echivalent în vechiul sistem cu 1,43 n mesaje sosite la controlorul de zboruri, această relație poate fi scrisă și sub forma

$$\text{procentul de mesaje eronate} = f = 0{,}015\ \ x, \tag{14.2}$$

unde x este procentul de ocupare a controlorului pentru rezolvarea mesajelor.

O relație de acest tip este analoagă cu rezultatele cunoscute din teoria informației și, mai cu seamă, în psihologie. Experimentele făcute au arătat că creșterea erorilor este proporțională cu creșterea nivelului de activitate. Dar ele au condus de asemenea la conceptul de volum de înregistrare ; acest concept a apărut mai semnificativ, datorită faptului că cantitatea de informații ce trebuie stocată în memoria controlorului la fiecare moment depinde de numărul de informații. Astfel, dacă ar exista un volum maxim de înregistrări fără erori atunci depășirea acestui volum înseamnă tocmai un număr ridicat de informații sosite la controlorul de zbor.

Totuși a fost necesar să legăm frecvența erorilor de cantitatea de informații stocate în memoria pe termen scurt și în acest scop a fost necesar să obținem numărul de informații memorate. S-a observat că timpul de stocare a unei informații a fost repartizat aproximativ exponențial cu media de 4,71 min. Prin urmare, creșterea și descreșterea volumului de informații memorate poate fi privită ca un proces

de naştere şi moarte, după cum urmează : 1) intensitatea naşterilor nulă ; 2) imigraţia : aleatoare, intrări independente cu media ν pe minut ; 3) intensitatea de moarte : aleatoare, independentă cu media μN, unde mărimea populaţiei este N.

Cititorul va remarca că acest model de naştere şi moarte este un caz particular al celor mai simple modele de aşteptare, cu mai multe staţii de servire, anume este $M/M/\infty$: (∞/FIFO) — un model cu un număr infinit de staţii de servire. În starea stabilă, repartiţia numărului de clienţi existenţi în sistem este Poisson, cu media ν/μ. (În momentul în care am întreprins cercetările, nu am gîndit în termenii unor modele cu o infinitate de staţii de servire şi am obţinut soluţia prin alte metode).

Trebuia acum să determinăm un standard de toleranţă pentru erori. S-a hotărît, destul de arbitrar, ca nivelul tolerabil al erorilor să fie 1%; aceasta corespundea unui număr de două informaţii pe minut memorate pe termen scurt. Nu toate mesajele erau independente între ele, unele avînd acelaşi conţinut, iar altele servind pentru precizarea unor mesaje anterioare. De exemplu, a apărut o serie de 14 mesaje toate în legătură cu un acelaşi incident petrecut cu un avion în curs de decolare.

S-a estimat că în medie ar fi aproximativ 1,6 informaţii cu acelaşi conţinut în memorie Numărul de conţinuturi memorate, corespunzînd unui nivel de eroare admis de 1%, a fost în medie 5,88. Aceasta poate fi interpretată astfel : capacitatea efectivă a memoriei controlorului de zbor era de aproximativ şase conţinuturi simultan. Evident că uneori se poate întîmpla ca memoria să conţină 10 sau 11 conţinuturi simultan. Aceasta a fost o metodă foarte aproximativă, empirică, de determinare a cantităţii de informaţie ce urmează a fi stocată în memoria controlorului de zboruri. Ea reprezintă un ghid pentru deciziile ulterioare.

Fără îndoială că, din punctul de vedere al exactităţii, controlorul de zbor era supraîncărcat în perioadele de vîrf, atîta timp cît două informaţii pe minut în memoria pe termen scurt este echivalent cu 2,86 pe minut în total, şi aceasta corespunzînd unei ocupări de circa 65% a controlorului de zbor cu prelucrarea acestor informaţii. În condiţiile noului sistem, cu un număr mai redus de informaţii de prelucrat, situaţia a fost uşor îmbunătăţită. A apărut însă evident că aceasta nu reprezenta decît o soluţie de compromis. Analizele statistice privind frecvenţa mesajelor au arătat că, în condiţiile noului sistem, numărul de mesaje prelucrate pe zbor este proporţional

cu numărul zborurilor ce urmează să decoleze în ora următoare și egal cu 9,5 mesaje pe zbor (repartiția nu a fost binomială). Se știa că numărul avioanelor ce urmau să decoleze în ora de vîrf de dimineața, în perioada verii anului 1960, era de aproximativ 22, astfel încît timpul mediu necesar a prelucra mesajele în această oră de vîrf era de $22 \times 9{,}5 \times 0{,}23 = 48{,}1$ min. Aceasta ar reprezenta o ocupare medie de circa 80%, în întregime format din mesaje ce necesită luarea unei decizii. O singură persoană nu ar putea face față acestei munci, de aceea ar trebui făcută o împărțire a ei între două persoane. Am fost puși în fața unei probleme speciale : să ne imaginăm un mijloc de reducere a informațiilor stocate pe termen scurt și un mijloc de a stoca aceste informații astfel încît doi controlori de zbor să poată avea acces la ele simultan. În mod experimental am imaginat un display* care era un mijloc foarte eficient de îmbunătățire a sistemului de memorare pe termen lung și în același timp capabil să rețină și informațiile pe termen scurt. Se părea însă că nu exista nici o modalitate de prezentare a datelor pe termen scurt care să permită celor doi controlori să le citească și să modifice fără un sistem de intercomunicare destul de complicat. Analizele făcute asupra datelor pe termen scurt colectate pe parcursul primelor faze ale studiului au relevat faptul că două persoane vor reuși doar să lucreze la 1,6 ori capacitatea unuia singur. Prin urmare, în 1960 intensitatea efectivă a traficului pentru fiecare controlor de zbor nu va fi 0,40, ci 0,50. Graficele pentru $M/M/2 : (\infty/\mathrm{SIRO})$ arată că pentru $\rho = 0{,}5$ aproximativ un mesaj din 12 va fi prelucrat întîrziat cu mai mult de 0,5 min.

Această valoare a fost considerată ca neacceptabilă. Devenise evident că se impunea o divizare a responsabilităților între cei doi controlori de zbor, care să le reducă necesitățile de intercomunicare. O asemenea diviziune a fost stabilită, odată cu un sistem de vizualizare a informațiilor. Teoria așteptării nu a fost utilizată în aceste faze finale ale sistemului, așa încît nu le vom acorda atenție.

Ar fi poate interesant de menționat faptul că sistemul de control rezultat din acest studiu a rămas neschimbat pînă în 1961. Atunci au fost efectuate o serie de modificări minore, care au avut ca rezultat o creștere a capacității. În momentul cînd scriu această carte, sistemul astfel modificat continuă să funcționeze satisfăcător.

---

* Dispozitiv de vizualizare a rezultatelor pe un ecran, asemănător televizorului. (*N. T.*)

15

O PROBLEMĂ A STANDARDELOR DE SERVIRE

Oficiile pentru rezervarea locurilor pentru cursele aeriene reprezintă primul contact direct al companiei cu posibilii clienţi ; de aceea trebuie să se asigure standarde minime de servire. Aceste standarde sînt de două tipuri : acelea care sînt legate de modul în care angajaţii, răspunzînd la apelurile telefonice, ascultă şi înregistrează dorinţele clienţilor şi cele legate de întîrzierile posibile ce pot să apară în încercarea clienţilor de a vorbi cu un angajat. Acestea din urmă pot fi măsurate suficient de corect şi sînt determinate aproape în întregime de numărul agenţilor şi de natura instalaţiei telefonice. Atîta timp cît sînt implicate oficiile de rezervări telefonice de locuri, două standarde sînt relevante ; anume cel care se referă la proporţia apelurilor respinse datorită blocării tuturor liniilor telefonice şi cel care se referă la proporţia apelurilor ce trebuie să aştepte mai mult decît o perioadă specificată de timp. Ceea ce se găseşte sub controlul direct al companiei este numărul de linii telefonice instalate şi numărul de angajaţi la oficiile de rezervare a locurilor, care le manevrează. În 1961, conducătorul departamentului de analiză a încasărilor de la Compania Air Canada i-a cerut lui H.J.G. Whitton de la departamentul de cercetări operaţionale să realizeze un studiu asupra relaţiilor existente între numărul de linii telefonice şi numărul angajaţilor, pe de o parte, şi frecvenţa apelurilor respinse şi repartiţia timpului de aşteptare, pe de altă parte. Studiul trebuia să cuprindă toate oficiile de rezervări de locuri ale Companiei Air Canada şi trebuia să ia în considerare o mare varietate a frecvenţelor de intrare şi a

repartiţiilor timpului de servire. Obiectivele acestui studiu erau următoarele :

1. Să permită conducătorilor de oficii de rezervări de locuri să estimeze mai precis numărul necesar de linii şi de angajaţi, pentru a face faţă volumului de apeluri de la oficiile lor, menţinînd un standard de servire corespunzător, măsurat de numărul apelurilor care aşteaptă.

2. Să obţină din rezultatele acestei prime faze o tehnică de calcul a cerinţelor totale de forţă de muncă pentru oficiile de rezervări de locuri, ţinînd cont de schimburile agenţilor, de concedii, zile de odihnă etc.

3. Să determine cea mai bună organizare în cadrul oficiilor de rezervare de locuri.

15.1. PRIMA FAZĂ

Primul lucru pe care l-a făcut Whitton a fost să observe operaţiile efectuate în cadrul unor oficii, pentru a aprecia modul de organizare. În această apreciere făcută, el scrie :

„Un sistem de rezervare de locuri constă dintr-o combinaţie a cîtorva procese simple fundamentale, fiecare putînd fi rezumat astfel. În momentul în care un client face apelul, el poate fi admis în sistem doar dacă una din liniile telefonice este liberă : dacă este admis, el vorbeşte cu un angajat, fie printr-un dispozitiv automat, fie printr-un procedeu de comutare manuală (în acest caz apelul va trebui să aştepte pînă cînd operatorul îl comută). În orice caz, apelul trebuie să aştepte pînă cînd un angajat este capabil să îl preia. Atunci cînd conversaţia a luat sfîrşit, linia ocupată este eliberată automat dar angajatul poate cere o perioadă suplimentară, pentru a-şi completa înregistrările, etc. Sistemele studiate constau dintr-un număr de astfel de sisteme simple, cu diferite tipuri de apeluri, recepţionate prin diferite combinaţii de linii, fiind dispecerizate în sistem printr-una din cele două metode mai importante. Alternativele posibile pentru fiecare fază a sistemului sînt prezentate în tabel.

Tabelul la care se face referirea este în această carte tabelul 15.1.

Whitton continuă :

Majoritatea oficiilor primesc apeluri de tipurile prezentate în prima coloană a tabelului, dar în timp ce la anumite oficii mici toate

rezervările pot fi recepționate la un singur grup de linii telefonice, în oficiile mai mari pot să existe grupuri separate pentru public, pentru intermediari, pentru apeluri speciale, sau pentru informații

Tabelul 15.1

Diferite tipuri de apeluri și angajați și metodele posibile de repartizare a apelurilor

Tipuri de apeluri recepționate	Grupuri posibile de linii	Metode de repartiție a apelurilor către angajați	Tipuri de angajați
Apeluri de rezervări (i) de la public	Linii publice de rezervări	Tablou de distribuție manual	Pentru rezervări interne
(ii) de la intermediari	Linii speciale pentru intermediari	Distribuitor automat	Pentru rezervări internaționale
(iii) surse speciale	Linii pentru reconfirmare	Direct la angajați (în oficiile mici)	Pentru apelurile intermediarilor
Apeluri de reconfirmare	Linii pentru anunțarea sosirilor avioanelor		Pentru diferite apeluri speciale
Apeluri pentru informații legate de orarul avioanelor	Linii administrative		Pentru apelurile legate de sosirile avioanelor
Apeluri administrative			

legate de sosirea avioanelor. Liniile pentru reconfirmare pot sau nu să fie separate. Angajații din oficiu pot fi grupați în diferite moduri — poate să existe un grup de angajați atît în apelurile internaționale cît și în cele ale intermediarilor (sau în apeluri speciale), sau pot

exista angajaţi separaţi, doar pentru apelurile internaţionale. Gruparea angajaţilor are un rol important în eficacitatea oficiilor (tratate independent) datorită faptului că în perioadele de vîrf angajaţii unui grup pot să se suplinească unul pe altul".

Următorul pas era de a analiza datele ce erau disponibile. Posibilităţile de a strînge date speciale erau foarte limitate, în schimb era disponibilă o mare cantitate de informaţie înregistrată la oficiile de rezervare a locurilor. Aceasta includea :

1. Numărul de apeluri acceptate pe oră, din categoria rezervărilor publice.

2. Înregistrări ale duratelor convorbirilor clienţilor cu operatorii de la tabloul de distribuţie, adică timpii de servire ai operatorilor.

3. Înregistrări ale duratelor convorbirilor clienţilor cu angajaţii de la oficiul de rezervări : timpii de servire ai angajaţilor.

În plus s-au făcut măsurări asupra :

4. Timpilor necesari angajaţilor pentru înregistrări, după încheierea convorbirilor cu clienţii.

Frecvenţele orare ale apelurilor posedă o caracteristică clară, anume modul de repartiţie a lor în cursul unei zile.

Figura 15.1 arată un astfel de mod de repartiţie, care este aproximativ acelaşi în toate zilele săptămînii, cu excepţia dispariţiei celor

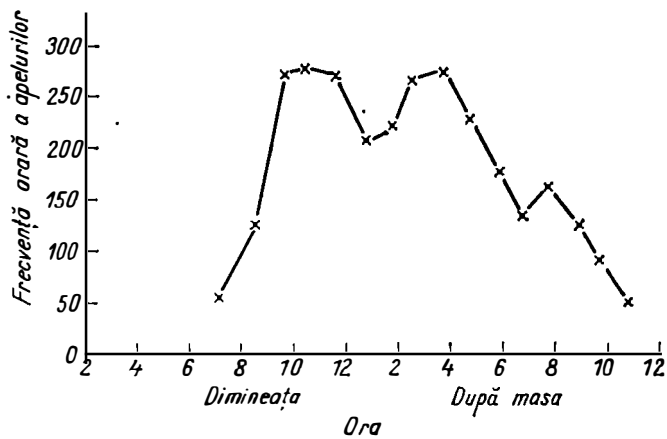


Fig. 15.1. Variaţia frecvenţei orare a apelurilor într-o zi tipică (în luna mai 1959).

două vîrfuri în zilele de sîmbătă și duminică. În plus, există anumite variații ale numărului de apeluri de la o zi la alta, în cursul săptămînii, și în plus există acele tendințe anuale și sezoniere care au constituit de multă vreme sursa principală a cîștigurilor companiilor aeriene.

Existența acestor variații înseamnă că este practic imposibil să găsim ore reprezentative, ce ar putea fi considerate ca selecții din aceeași populație. Datele nu pot fi folosite pentru a indica dacă repartiția apelurilor este sau nu Poisson. Nu se pot efectua observații directe asupra intervalelor dintre apeluri, astfel încît a trebuit să se considere că repartiția acestor intervale este exponențială (așa cum se obișnuiește dealtfel în telefonie).

Într-un oficiu, în care era un tablou de distribuire manual, timpul de manevrare a unui apel de către operator are o repartiție lognormală cu vîrful la 4 s și media de 6 s. Observațiile au arătat că diversele tipuri de apeluri aveau durate diferite, variind în medie de la 90 la 200 de secunde. Formele repartițiilor erau foarte asemănătoare

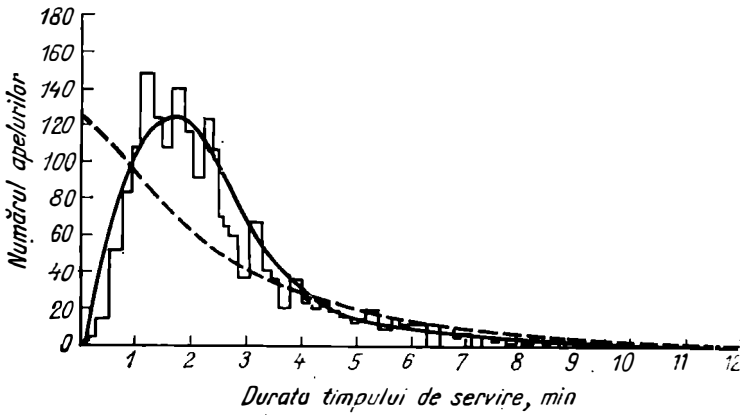


Fig. 15.2. Repartiția timpilor de servire ai angajaților unui oficiu de rezervări de locuri pentru public: — — — exponențială; ·——— lognormală.

în jurul mediilor. Figura 15.2 prezintă una dintre aceste repartiții, împreună cu repartițiile exponențială și lognormală corespunzătoare datelor observate.

Whitton a ajustat datele şi cu o repartiţie Pearson de tip III, dar aceasta, care nu este reprezentată în figură, era apropiată de cea exponenţială. Deşi repartiţia lognormală părea că aproximează bine datele observate, totuşi testul χ^2 a contrazis acest lucru ($\chi^2 = 83{,}08$ cu 44 grade de libertate). Totuşi, probabil că este destul de dificil să găsim o formă mai bună şi în plus trebuie să ţinem seamă de particularităţile muncii fiecărui angajat. Aşa că am folosit repartiţia lognormală în calculele ulterioare fără prea multe ezitări.

În 80% din cazuri, angajaţii nu au nevoie de timp suplimentar la sfîrşitul unei convorbiri, dar cînd există şi astfel de situaţii, timpii corespunzători sînt repartizaţi din nou lognormal cu media 1,75 min.

La acest punct, Whitton a luat în considerare diferitele modalităţi în care putea analiza diversele sisteme de aşteptare complexe cu care era confruntat. El arată:

„Caracterul neexponenţial al repartiţiei timpilor de servire şi lipsa unor date despre repartiţia apelurilor înseamnă că orice rezultat dedus din teoria exponenţială a aşteptării va fi îndoielnic. Această observaţie, cuplată cu faptul că ecuaţiile ar fi extrem de complicate pentru orice versiune nesimplificată a sistemului, ne conduce la concluzia folosirii simulării cu ajutorul unui calculator electronic. Aceasta nu înseamnă că posibilitatea obţinerii unor rezultate teoretice este respinsă în întregime; se speră că aproximări satisfăcătoare bazate pe teorie să fie în continuare posibile, dar se consideră că ele nu ar fi putut fi obţinute rapid şi utilizabil. Rezultatele numerice cerute puteau fi obţinute printr-o simulare relativ simplă, dar problemele legate de dirijarea diferitelor tipuri de apeluri etc. puteau fi studiate doar dacă simularea reproducea funcţiile esenţiale ale sistemului real. De aceea, simularea a fost făcută cît mai aproape de realitate."

15.1.1. SIMULAREA

În descrierea simulării oficiilor de rezervare de locuri Whitton a scris următoarele:

„Metoda simulării se bazează pe faptul că evoluţia unui apel printr-un sistem real poate fi vizualizată ca o serie de mutări dintr-o stare în alta, apelul fiind în fiecare stare un interval finit de timp.

Dacă un apel este acceptat, acesta ajunge la tabelul de distribuţie şi intră într-una din cele două stări — este servit imediat sau trebuie să aştepte, dar în fiecare caz trece un interval de timp înaintea schimbării stării sale. Aşadar, simularea constă într-o procedură care să ne asigure că modificările de stare a tuturor apelurilor din sistem sînt corect sincronizate : modificările de stare sînt însoţite de secvenţe logice, proiectate pentru a corespunde, pe cît posibil, operaţiei din sistemul real.

Metoda de sincronizare folosită este o trăsătură de bază a simulării avînd un efect major asupra timpului de calcul necesar obţinerii rezultatelor. Metoda cea mai cunoscută presupune verificarea fiecărui apel la sfîrşitul fiecărei unităţi de timp simulate, pentru a determina ce modificări de stare au apărut în intervalul considerat : în cazul nostru intervalele între modificările de stare erau relativ lungi comparativ cu unitatea de timp utilizată şi astfel s-a considerat că simularea ar dura mai puţin dacă s-ar realiza pe baza trecerii de la o stare de alta (adică să se utilizeze metoda „duratei minime"). Momentele la care apar modificări de stare sînt reţinute în memorie în ordine cronologică, iar atunci cînd procesul de simulare a ajuns la sfîrşitul unei secvenţe logice, contorul de timp este deplasat la momentul următoarei modificări de stare înregistrate şi o nouă secvenţă logică va începe. Acest proces este ilustrat în fig. 15.3, care prezintă forma simplificată a schemei logice ce apare în procesul simulării sistemului de rezervări telefonice pentru zboruri interne. Pentru simplificare, schema logică exclude toţi paşii care presupun ajustări şi nu se arată conexiunile cu celelalte părţi ale întregului sistem. Oricum, se arată modul în care toate secvenţele logice trebuie să conducă la alegerea următoarei modificări de stare şi deci la următoarea secvenţă.

În capitolul referitor la tehnicile de simulare am analizat pe scurt trei operaţii necesare în simularea modelelor de aşteptare : generarea numerelor aleatoare, producerea sosirilor clienţilor şi a timpilor de servire. Remarcăm că în multe cazuri nu sînt necesare numere aleatoare veritabile. Acest lucru s-a întîmplat şi în cercetările întreprinse de Whitton. Legat de aceste trei aspecte majore ale simulării el spune :

„Numerele aleatoare sînt necesare foarte frecvent în simulare şi ele sînt generate de un proces multiplicativ congruenţial. Pentru acest

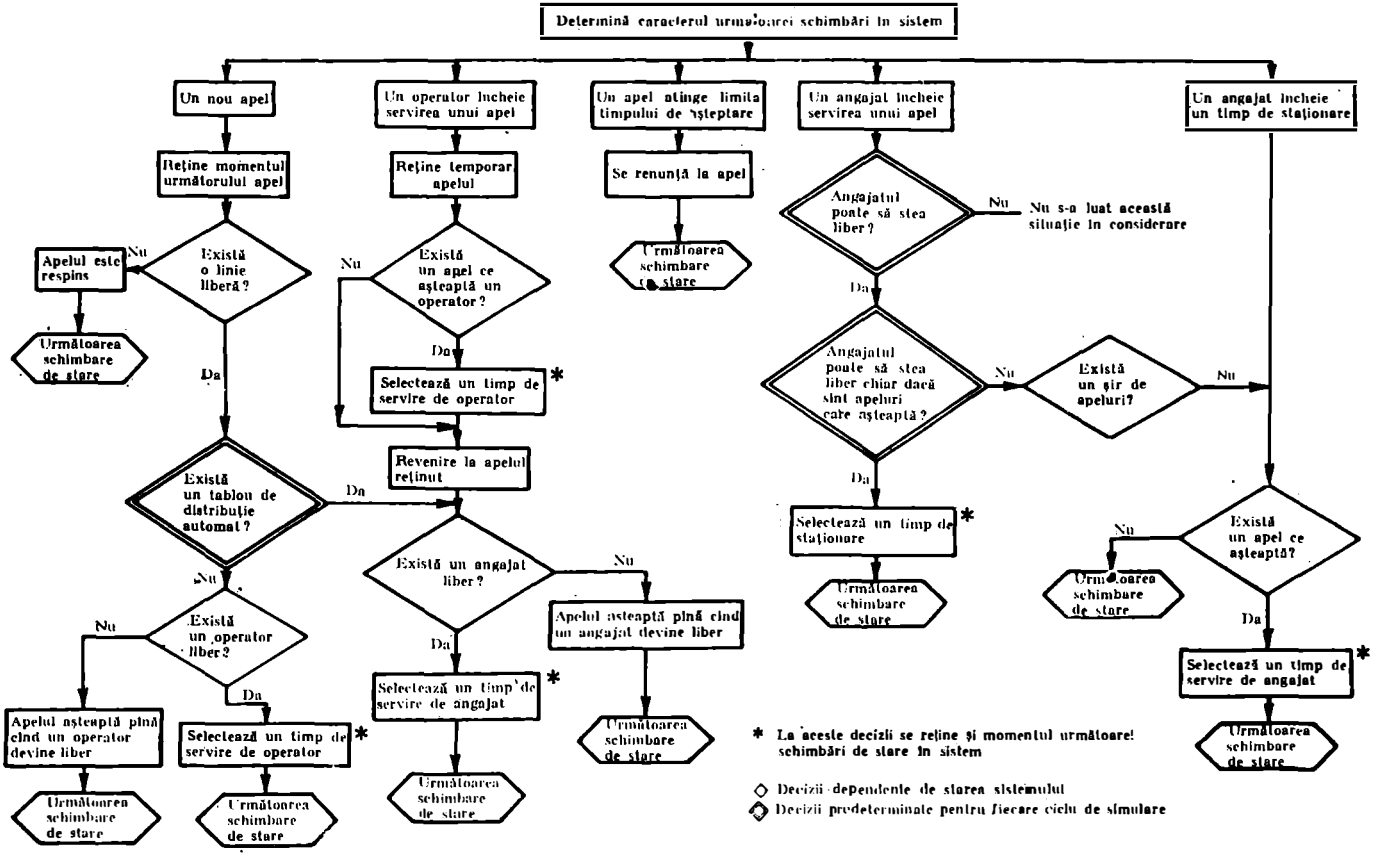

Determină caracterul următoarei schimbări în sistem
Un nou apel
Reține momentul următorului apel
Există o linie liberă?
Nu
Apelul este respins
Următoarea schimbare de stare
Da
Există un tablou de distribuție automat?
Nu
Există un operator liber?
Nu
Apelul așteaptă pînă cînd un operator devine liber
Următoarea schimbare de stare
Da
Selectează un timp de servire de operator *
Următoarea schimbare de stare
Un operator încheie servirea unui apel
Reține temporar apelul
Există un apel ce așteaptă un operator?
Nu
Da
Selectează un timp de servire de operator *
Revenire la apelul reținut
Da
Există un angajat liber?
Nu
Apelul așteaptă pînă cînd un angajat devine liber
Următoarea schimbare de stare
Da
Selectează un timp de servire de angajat *
Următoarea schimbare de stare
Un apel atinge limita timpului de așteptare
Se renunță la apel
Următoarea schimbare de stare
Un angajat încheie servirea unui apel
Angajatul poate să stea liber?
Nu
Nu s-a luat această situație în considerare
Da
Angajatul poate să stea liber chiar dacă sînt apeluri care așteaptă?
Nu
Există un șir de apeluri?
Nu
Da
Selectează un timp de staționare *
Următoarea schimbare de stare
Un angajat încheie un timp de staționare
Există un apel ce așteaptă?
Nu
Următoarea schimbare de stare
Da
Selectează un timp de servire de angajat *
Următoarea schimbare de stare
* La aceste decizii se reține și momentul următoarei schimbări de stare în sistem
◇ Decizii dependente de starea sistemului
◇ Decizii predeterminate pentru fiecare ciclu de simulare

Fig. 15.3. Versiune simplificată a schemei logice pentru simularea prelucrării apelurilor de către angajați.

tip de proces este ales un multiplicator p, iar numerele aleatoare succesive sînt obținute din relația

$$x_{n+1} = px_n(\text{mod } P),$$

unde x_0 este numărul aleator inițial. În cazul nostru multiplicatorul a fost ales 3^{17} (adică de forma 3^{4k+1}), iar P a fost luat 8. Acest multiplicator dă cicluri de lungime maximă și va produce 5×10^6 numere distincte înaintea de primă repetare (este puțin probabil ca să fie utilizate mai mult de 100 000 numere aleatoare, într-un ciclu normal de simulare). Se presupune că apelurile sosesc la intervale exponențiale, care pot fi generate de procedura realizată de von Neumann, descrisă în cap. 7. Timpii de servire sînt obținuți într-o manieră convențională. Se memorează un tabel al funcției de repartiție $F(z)$ a acestei variabile aleatoare z; se consideră un număr aleator x și se caută în tabel cea mai mică valoare z pentru care $F(z) \geqslant x$. Dacă funcția de repartiție $F(z)$ ar fi avut o formă mai prelucrabilă decît longnormală, atunci am fi putut folosi o metodă similară cu cea prezentată în cap. 6".

Intrările simulate în oficiul de rezervare de locuri erau de două tipuri: valori numerice date parametrilor și indicative care codifică organizarea oficiului studiat, în concordanță cu o anumită clasificare. Aceste intrări pot fi rezumate în următoarea listă. Valorile numerice cerute erau:

1. Frecvențele orare ale:
apelurilor pentru rezervări de locuri pe liniile telefonice pentru public;
apelurilor pentru rezervări sosite pe alte linii;
apelurilor de reconfirmare;
apelurilor de la intermediari;
apelurilor speciale;
apelurilor pentru comunicarea de informații legate de sosirile avioanelor;
apelurilor administrative.

2. Numărul de linii pentru:
rezervări telefonice pentru public;
rezervări pentru alte categorii;
reconfirmare;
apeluri de la intermediari;

apeluri speciale;
apeluri administrative.

3. Numărul de operatori şi repartiţia timpului de servire a operatorilor la tabelul de distribuţie.

4. Numărul de angajaţi necesari pentru:
apeluri interne de rezervări;
apeluri interne şi internaţionale;
apeluri de la intermediari;
apeluri speciale;
apeluri pentru informaţii legate de orarul avioanelor, împreună cu repartiţiile timpilor de servire corespunzători.

5. Repartiţia timpilor necesari angajaţilor pentru a face înregistrări etc.

6. Timpul maxim de aşteptare permis (dacă se cere).

7. Timpul standard de aşteptare.

8. Numărul aleator iniţial şi multiplicatorul cerut.

9. Numărul de apeluri pentru rezervări odată cu care se va termina simularea.

Chestiunile de organizare a oficiului erau:

10. Există un tablou de distribuţie automat?

Apelurile de la intermediari sau cele speciale pot ajunge direct la angajaţi?

Există linii speciale pentru apelurile pentru informaţii legate de sosirile avioanelor?

Aceste apeluri ajung direct la angajaţi?

Apelurile internaţionale sînt rezolvate de către angajaţi sau de angajaţi speciali?

Se acordă un interval de timp pentru transferul unui apel?

Un angajat îşi poate permite un timp suplimentar, după încheierea unei convorbiri?

Există un timp maxim de aşteptare?

Ieşirile produse prin simulare, după analiza numărului de apeluri prevăzute iniţial, iau următoarea formă:

Numărul de apeluri recepţionate şi numărul de apeluri respinse	pentru fiecare tip de apel
Procentajul de utilizare	pentru fiecare grup de linii pentru operatori pentru fiecare tip de angajat

Lungimea medie şi maximă a şirului, pentru fiecare staţie de servire :
Procentul de apeluri ce aşteaptă mai mult decît timpul standard } pentru fiecare tip de apel

15.1.2. APLICAŢIE A SIMULĂRII

După cum am văzut în cap. 6, simulările pot da forme variate de distorsionări. Cauzele acestor devieri erau posibilele erori în reprezentarea logică a sistemului şi insuficienta atenţie acordată culegerii datelor. Ambele pot produce perturbaţii şi imprecizie. În plus, imprecizia poate fi, ca de obicei, cauzată de caracterul aleator al procesului de simulare.

Cea mai cunoscută metodă de control al perturbaţiilor este de a compara rezultatele ciclurilor de simulare cu observaţiile operaţiilor reale. Fără îndoială, aceasta este greu de realizat. Atîta timp cît variabilitatea şi imprecizia constituiau problemele majore, era mai întîi necesar ca fiecare ciclu de simulare să fie suficient de lung pentru a asigura ca rezultatele finale să corespundă stării stabile.

Această problemă a fost cercetată prin realizarea unei serii de teste, în care programul a fost modificat astfel încît a afişat rezultatele cumulate, la sfîrşitul fiecăror 100 de apeluri recepţionate simulate. Unele din rezultatele obţinute sînt prezentate în fig. 15.4,

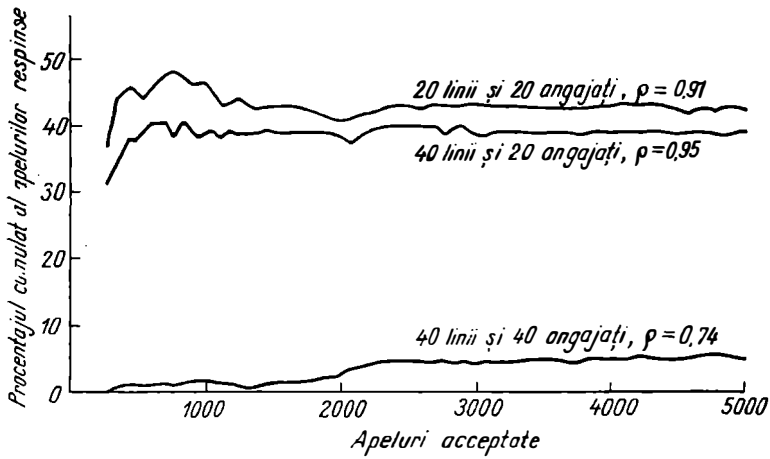


Fig. 15.4. Aspectul progresiv al procentului apelurilor respinse (intensitatea intrărilor = 600 apeluri/h).

unde procentajul cumulat al apelurilor respinse este expus comparativ cu numărul de apeluri recepţionate într-un sistem în care apelurile străbat tabloul de distribuţie pentru a ajunge la angajaţi (adică cu două faze de servire, în serie).

Testele au arătat că procentul de apeluri respinse atinge regiunea valorii stării stabile după prelucrarea a circa 1000 de apeluri simulate, dar fluctuaţiile în jurul acestei valori se menţin pînă la circa 3000 de apeluri. Modificările ce apar pînă la atingerea valorii stării stabile nu urmează un tipar coerent. Probabilitatea unei întîrzieri avea cam acelaşi aspect în starea de tranziţie atunci cînd numărul de linii era identic cu numărul angajaţilor, dar atunci cînd există mai multe linii decît angajaţi, probabilitatea unei întîrzieri se modifică foarte lent şi exponenţial spre valoarea stării stabile.

Pentru a asigura un grad ridicat de precizie aproximaţiilor valorilor stării stabile, ciclurile de simulare au fost continuate pentru cel puţin 3000 de apeluri.

Tabelul 15.2 prezintă mărimea variaţiilor aleatoare la care ne putem aştepta, rezultate obţinute prin simulare. Pentru fiecare situaţie, ultimele coloane conţin procentul de apeluri respinse obţinut

Tabelul 15.2

Variaţiile aleatoare ale procentului apelurilor respinse

Metoda de distribuire a apelurilor	Numărul angajaţilor	Apelurile respinse %		
Distribuitor automat	20	36,2	37,5	37,0
	40	1,8	1,2	2,2
Tablou de distribuţie manual	20	38,3	40,5	37.0
	40	2,2	3,0	1,8

în trei cicluri de simulare, care diferă doar prin alegerea numărului aleator iniţial. Toate rezultatele se referă la o frecvenţă a intrărilor de 600 de apeluri pe oră, 40 de linii telefonice, un timp mediu de servire al unui angajat de 3 min; în cazul ultimelor două situaţii se consideră un tablou de distribuţie manual cu trei operatori, avînd un timp mediu de servire de 6 s.

Aceste variaţii sînt suficient de mici comparativ cu modificările ce apar în cazul unor alterări ale parametrilor de intrare (de exemplu

ale numărului de linii sau de angajaţi), şi astfel un număr mare de cicluri de simulare nu şi-ar avea rostul decît în cazuri cu totul excepţionale.

În zona de interes (aproximativ 15 % de apeluri respinse), la un nivel de semnificaţie de 95 %, limitele de încredere ale rezultatului unui singur ciclu de simulare vor fi de aproximativ $\pm$ 2 %.

Variaţia procentului apelurilor care aşteaptă mai mult decît timpul standard este de aceeaşi mărime.

15.1.3. REZULTATE OBŢINUTE PRIN SIMULARE

Pentru diferite valori ale frecvenţelor intrărilor, ale numărului de linii telefonice, de angajaţi etc., prin simulare se produc valori ale procentului de apeluri respinse şi procentului de apeluri ce aşteaptă o perioadă mai mare decît limita specificată. Pentru a estima numărul de linii şi angajaţi necesari, a asigura o servire standard pentru orice frecvenţe ale intrărilor, primul pas a fost realizarea unei serii de cicluri de simulare care să acopere regiunea de interes. Rezultatele acestor cicluri de simulare sînt prezentate în fig. 15.5 şi se referă la un anumit oficiu în care apelurile pentru rezervări străbăteau un tablou manual de distribuţie pentru a ajunge la angajaţi. Curbele sînt specifice sistemelor de aşteptare cu mai multe staţii de servire; ele arată că există de obicei o regiune critică legată de numărul de angajaţi (standardul specificat de companie părea să cadă în această regiune). Familia de curbe, pentru valori ridicate ale intrărilor, prezintă efectul limitării dimensiunilor şirului de aşteptare (adică a numărului de linii telefonice). Dacă numărul de linii telefonice creşte, procentul de apeluri respinse va scădea şi curbele vor avea un aspect aparte; dacă numărul de angajaţi nu va creşte, atunci timpii de aşteptare ai apelurilor acceptate vor fi mai lungi.

Au fost realizate şi curbele pentru procentul de apeluri respinse. Acestea arată că (pentru un număr fixat de linii), odată cu creşterea frecvenţei intrărilor, devine imposibil să se reducă procentul de apeluri respinse sub o anumită valoare. Dacă această valoare este mai ridicată decît standardul, singura modalitate de a atinge standardul este de a creşte numărul de linii telefonice şi deci numărul de angajaţi.

Estimaţii privind numărul de linii şi de angajaţi, necesari în diferite oficii de rezervare, pot fi obţinute în mod asemănător. S-a remar-

cat totuşi, că tabelele au o utilitate practică mai mare decît o au graficele de tipul celui din fig. 15.5, de aceea s-a hotărît completarea unor tabele, ce vor fi larg răspîndite.

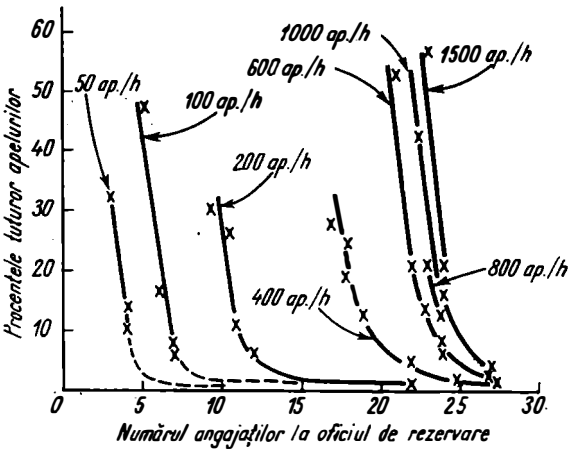


Fig 15.5. Procentele tuturor apelurilor de rezervare ce aşteaptă mai mult de 0,5 min pentru diverse frecvenţe de apeluri şi numere de angajaţi (rezultate obţinute prin simulare).

Tabelele vor da şi pentru cazurile în care rezultatele simulării arată o utilizare a angajaţilor în proporţie $\geqslant$ 70 %, numărul de angajaţi necesari pentru o utilizare mai puţin obositoare.

O variabilă pe care nu am discutat-o încă este timpul mediu de servire al angajaţilor. Acesta poate să difere de la un oficiu la altul, dar a trebuit să fie conţinut în tabele. Efectul modificărilor acestei variabile a fost punctul central al cercetărilor efectuate în legătură cu introducerea sistemului electronic de rezervare. S-a sugerat că introducerea acestui sistem va afecta timpul necesar unui angajat pentru a trata cu clientul. Aceasta ar fi putut fi adevărat, astfel că problema a fost cercetată iar rezultatele acestor cercetări, realizate la un oficiu cu tablou de distribuţie automat, sînt prezentate în tabelul 15.3.

Tabelul 15.3

Efectul variaţiei timpilor medii de servire asupra numărului de linii şi angajaţi

Timpul mediu de servire (min)	1,5	2	3	4
Numărul necesar de linii în orele de vîrf	19	24	33	40
Apeluri pe oră	Numărul angajaţilor			
200	10	12	15	17
400	20	25	33	41
600	24	33	45	56

Numărul de linii a fost astfel ales, încît procentul de apeluri respinse să nu depăşească standardul în orele de vîrf; acelaşi număr de linii a fost utilizat pentru două frecvenţe mai scăzute ale intrărilor. Numărul angajaţilor a fost dat de rezultatele simulărilor, ajustate pentru a înlătura o utilizare mai mare de 70%.

15.1.4. COMPARAŢII CU TEORIA AŞTEPTĂRII

Whitton a tras concluzia, corectă dealtfel, că o continuare ce constă dintr-un număr mare de cicluri de simulare, producînd volume imense de date, nu era cea mai bună soluţie. El a considerat că dacă se putea găsi o cale de utilizare a modelelor teoretice existente pentru a produce aceleaşi rezultate ar fi mult mai bine chiar dacă de fapt aceste modele diferă foarte mult în structura lor de ceea ce se întîmplă în oficiile de rezervare de locuri. De aceea, el a realizat o comparare a rezultatelor obţinute în cîteva cicluri de simulare cu cele rezultate din unele modele matematice. Vom continua prezentarea acestui stadiu al primei faze cu propriile sale cuvinte.

„Cel mai simplu sistem de aşteptare cuprins în sistem este unul în care apelurile sosesc aleator şi sînt tratate în ordinea sosirii de către oricare dintre angajaţi, dimensiunea şirului de aşteptare fiind limitată de numărul fix de linii telefonice. Dacă timpii de servire sînt presupuşi a fi repartizaţi exponenţial, atunci pot fi obţinute valorile teoretice pentru diferite statistici utile, prin metode standard.

Valorile deduse dintr-un astfel de sistem vor fi valabile numai pentru tipul cel mai simplu de oficiu, în care apelurile ajung direct la angajaţi, fără a trece printr-o fază intermediară de aşteptare la

tabloul de distribuţie. Cele mai multe oficii însă au tablouri de distribuţie, astfel că cea mai evidentă generalizare a unui sistem simplu constă în introducerea unui al doilea stadiu de servire, adică următorul sistem ce trebuie luat în consideraţie este un sistem ce conţine două faze în serie, fiecare cu mai multe staţii de servire, cu o limitare superioară a numărului de apeluri acceptate în sistem. Formulele aplicabile unui astfel de sistem nu pot fi găsite în literatură, dar dificultăţile legate de deducerea lor sînt numai algebrice.

Cea mai facilă ilustrare a modului în care expresiile teoretice corespund rezultatelor simulării constă într-o comparaţie a estimaţiilor procentului de apeluri respinse. O astfel de comparaţie poate fi făcută cu ajutorul următorului tabel, care dă estimaţiile acestor procente pentru o frecvenţă a intrărilor de 600 apeluri pe oră, cu 40 de linii telefonice şi un timp mediu de servire de 3 min pentru sistemele cu o singură fază şi respectiv timpii medii de servire de 6 s şi respectiv 3 min pentru sistemele formate din două faze (tabelul 15.4). Diferen-

Tabelul 15.4

Comparaţie a estimaţiilor teoretice ale procentului de apeluri respinse, cu estimaţiile produse prin simulare

Numărul fazelor	Numărul angajaţilor		Estimarea teoretică	Estimarea prin simulare (media a trei cicluri)
1	20		33,3	36,9
	40		1,4	1,8
	Prima fază	A doua fază		
2	2	20	33,3	38,6
2	2	40	2,2	2,3

ţele ce apar în cele două cazuri cu 40 de linii şi 20 de angajaţi sînt prea mari pentru a fi atribuite în întregime variaţiilor aleatoare: acestea ar putea fi cauzate fie de caracterul nexeponenţial al repartiţiilor timpului de servire, fie de caracterul prea grosolan al teoriei. Repartiţiile teoretice ale timpului de aşteptare pentru cazul cu o singură fază au aceeaşi formă cu cele produse prin simulare (adică acelaşi grad de asimetrie) dar valorile medii ale celor două tipuri de repartiţie diferă. Şi aici probabil că o îmbunătăţire a teoriei ar putea furniza aproximaţii mai bune".

15.1.5. SFÎRŞITUL PRIMEI FAZE

În acest moment, Compania Air Canada posedă o simulare care va arăta, pentru diferite frecvenţe de intrare repartiţii ale timpului de servire şi regula FIFO a disciplinei de aşteptare, numărul de linii telefonice şi de angajaţi, procentul de apeluri respinse şi procentul de apeluri acceptate, care trebuie însă să aştepte mai mult decît un timp specificat. Simularea a fost programată pentru un calculator IBM 7070. În 1962 au fost simulate un număr mare de cicluri pentru a reprezenta unul din cele mai mari oficii de rezervări de locuri ale companiei. Acest sistem utilizează un tablou de distribuţie automată, cu acces direct, şi este completat cu circuite automate care asigură o perfectă regulă FIFO a disciplinei de aşteptare. În urma ciclurilor de simulare s-au completat cîteva tabele.

Imediat după ce aceste tabele au fost completate, s-a efectuat o reorganizare în cadrul companiei. Scopul acestor modificări a fost realizarea unei mai bune funcţionări în faţa cererii crescînde. Ca rezultat, Whitton, a cărui experienţă pentru sistemele de mari dimensiuni era considerabilă, a primit noi responsabilităţi, iar legătura sa cu proiectul pe care-l discutăm a fost întreruptă.

15.2. FAZA A DOUA

Sarcina continuării proiectului şi a colaborării cu departamentul de analiză a încasărilor i-a revenit lui R. W. Linder şi autorului. Este, cred, un fapt incontestabil — deşi regretabil şi neconvenabil — că realizarea unui proiect de cercetare operaţională depinde mai curînd de persoanele angajate direct decît de tehnicile care se pot utiliza. În cazul prezentat, era inevitabil ca modificările ce au survenit în echipa de studiu să influenţeze modul de abordare a problemei şi într-adevăr, aşa a şi fost.

Împreună cu Linder ne-am axat pe două probleme :

1. Standardele de servire pe care compania doreşte să le impună ca *minime* în întregul sistem.

2. Principalul rezultat al proiectului, anume calculul necesarului de forţă de muncă, cerut de modificările survenite, pentru oficiile de rezervare de locuri.

Prin aceasta, noi am privit lucrurile în cu totul alt mod.

Standardele minime ale companiei pentru rezolvarea apelurilor telefonice erau :

a) să nu fie respinse mai mult de 15% din apeluri, *în orice oră* ;

b) să nu fie mai mult de 18% apeluri acceptate, *în orice oră*, care să stea în aşteptare mai mult de 20 s, înainte de a fi preluate de către un angajat.

Tabelele obţinute prin simulare puteau fi imediat aplicate, pentru a obţine numărul de linii şi angajaţi necesar a satisface aceste condiţii, la diferite nivele de trafic. Exemple de astfel de tabele sînt reprezentate în tabelele 15.5 şi 15.6. Să analizăm aceste tabele, în scopul identificării unor reguli de lucru, uşor de aplicat de către conducătorii locali şi în acelaşi timp să găsim o metodă aproximativă pe care să o putem utiliza în alte studii, în locul unor cicluri de simulare costisitoare.

După o scurtă căutare, am stabilit că numărul minim de linii şi de angajaţi necesari asigurării standardului *minim* al companiei, poate fi calculat astfel :

1. Numărul necesar de angajaţi în fiecare oră este $E + 3$, unde E este numărul de erlangi ai traficului *acceptat* în ora respectivă, sau 85% din traficul oferit.

2. Numărul necesar de linii este $E + 5$.

Tabelul 15.7 are ca scop compararea numărului de linii şi de angajaţi necesari, rezultaţi în urma simulării şi în urma aplicării regulii „trei plus cinci", pentru a face faţă volumelor de trafic oferit, la nivelul standardelor minime ale companiei.

Comparaţia realizată arată abateri mici pentru valori scăzute ale nivelelor de trafic şi abateri nesemnificative pentru nivele ridicate. Ni s-a părut rezonabil să recomandăm utilizarea acestei reguli de aproximare de către conducătorii locali, în limita condiţiilor precizate şi pentru valori ale traficului dintre 5 şi 25 de erlangi (în afara acestor limite, se pare că regula nu este bună). Aceasta era scopul simulării : obţinerea unei reguli de aproximare pentru calculul necesarului de linii şi de angajaţi. În realitate, anumite oficii prezintă standarde de servire mai ridicate decît cele minime impuse de companie (un fenomen des întîlnit în transporturile aeriene) ; de aceea pentru aceste oficii este necesară determinarea unei alte reguli (aceasta a fost regula „patru plus şapte"). Fără îndoială, că prin rezolvarea acestor probleme s-a făcut un pas important în realizarea proiectului.

Tabelul 15.5

Procentul de apeluri respinse

Frecvența: 250 apeluri/oră Timp de servire: 4 min

Nr. de angajați									
22	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0
19	2	2	2	2	2	2	2	1	1
18	6	6	6	5	5	5	5	4	4
17	10	10	10	9	9	9	9	8	8
16	14	14	13	13	12	12	12	11	11
15	18	18	17	16	16	16	15	15	15
14	23	23	22	22	21	21	21	20	20
13	27	27	27	26	26	26	26	25	25
12	32	32	32	31	31	31	31	31	31
11	39	39	39	38	38	38	38	38	38
Nr. de linii	18	19	20	21	22	23	24	25	26

Frecvența: 400 apeluri/oră Timp de servire: 2,5 min

Nr. de angajați									
22	12	9	7	4	2	2	2	1	1
21	13	10	8	6	3	3	2	2	1
20	14	12	10	8	6	5	4	3	2
19	15	13	11	9	7	6	5	4	3
18	16	14	12	10	8	7	6	6	5
17	17	15	14	12	10	9	9	8	8
16	18	17	15	13	12	12	11	11	11
15	20	19	18	17	16	15	15	14	14
14	23	23	22	21	20	20	20	19	19
13	25	25	25	24	24	24	24	23	23
12	32	32	31	31	31	31	30	30	30
11	36	36	36	36	36	36	35	35	35
Nr. de linii	18	19	20	21	22	23	24	25	26

Frecvența: 500 apeluri/oră Timpul de servire: 2 min

Nr. de angajați									
22	6	5	5	4	4	4	3	2	1
21	7	7	6	5	5	5	4	3	2
20	9	8	7	6	6	5	5	4	3
19	10	9	8	7	7	6	5	4	4

Tabelul 15.5 (continuare)

Nr. de angajaţi									
18	11	10	9	8	8	7	7	6	6
17	13	12	11	10	10	9	9	8	8
16	15	15	14	13	13	12	12	11	11
15	19	18	17	16	15	15	15	15	15
14	22	21	21	20	20	20	20	20	20
13	26	25	24	23	23	23	23	23	22
12	32	32	31	31	31	31	31	31	31
11	38	38	38	37	37	37	37	37	37
Nr. de linii	18	19	20	21	22	23	24	25	26

Principala problemă era următoarea : ca să fie capabil să folosească aceste reguli, un conducător trebuia să poată să procure sau să estimeze volumul traficului pentru viitor. Cu alte cuvinte, el trebuia să prognozeze valori ale intensităţii intrărilor şi ale timpilor medii de servire. Acestea din urmă sînt constante în timp. Frecvenţele intrărilor însă sînt fluctuante în timpul zilei, săptămînal şi anual. Trebuia deci să rezolvăm într-un fel această problemă. După o examinare a înregistrărilor anterioare, am dezvoltat o metodă de previzionare a frecvenţelor orare ale apelurilor, deduse din previzionarea numărului de pasageri. Aceasta utiliza o relaţie între frecvenţele apelurilor la un oficiu de rezervări dintr-un anumit oraş şi numărul pasagerilor ce utilizau cursele aeriene. S-au obţinut astfel relaţii foarte utile.

O altă problemă pe care am întîlnit-o în această fază a studiului a fost aceea de estimare a numărului de clienţi ce cheamă telefonic oficiul de rezervări, în condiţiile în care singurele date înregistrate erau numărul de apeluri acceptate şi numărul celor respinse. O scurtă reflecţie a cititorului îl va convinge că prin adunarea acestora nu vom obţine numărul total de apeluri. Un client, al cărui apel este respins prima oară, în general nu renunţă. Aşa cum făcea d-na Overlander, el mai încearcă de cîteva ori. Fiecare încercare nereuşită se adună la numărul apelurilor respinse.

Considerînd acceptarea sau respingerea apelurilor la prima, la a doua şi la următoarele încercări, a fost posibil să obţinem o estimaţie brută a cererii adevărate, în fiecare oră. S-a presupus că un apel se

Tabelul 15.6

Procentul de apeluri acceptate, întîrziate mai mult de 20 s

Frecvența : 250 apeluri/oră Timpul de servire : 4 min

Nr. de angajați									
22	0	1	1	2	2	2	2	2	2
21	1	2	2	2	3	3	3	4	4
20	2	2	2	3	4	5	7	8	10
19	3	4	5	6	8	11	14	17	20
18	6	8	10	13	15	20	25	30	36
17	14	17	20	23	26	37	48	63	79
16	26	32	38	44	50	60	70	80	91
15	56	61	65	70	74	79	84	89	94
14	75	79	83	87	91	92	94	96	97
13	89	91	93	95	97	97	98	98	99
12	95	96	97	98	99	99	99	99	99
11	98	98	98	99	99	99	99	99	99
Nr. de linii	18	19	20	21	22	23	24	25	26

Frecvența : 400 apeluri/oră Timpul de servire 2,5 min

Nr. de angajați									
22	0	0	1	1	2	4	5	7	8
21	1	2	2	3	3	5	8	11	13
20	2	4	5	6	7	11	14	18	21
19	3	6	8	11	14	19	24	29	34
18	6	11	16	21	27	33	39	44	49
17	12	20	29	37	46	50	55	60	64
16	25	34	43	52	62	66	70	74	78
15	50	56	63	70	76	80	83	87	91
14	67	73	78	84	89	91	93	95	96
13	84	86	89	92	95	96	97	97	98
12	92	93	95	97	98	98	99	99	99
11	98	98	98	99	99	99	99	99	99
Nr. de linii	18	19	20	21	22	23	24	25	26

Frecvența : 500 apeluri/oră Timpul de servire : 2 min

Nr. de angajați									
22	0	0	1	1	2	3	5	6	8
21	0	1	2	4	5	7	9	11	13
20	0	2	4	6	8	11	14	18	21
19	1	4	7	11	14	19	24	29	34

Tabelul 15.6 (continuare)

Nr. de angajaţi									
18	3	9	14	20	25	31	37	44	50
17	9	18	26	35	43	48	53	58	63
16	19	29	40	50	61	65	69	73	77
15	37	47	57	67	76	80	83	87	90
14	59	66	74	82	89	90	92	94	96
13	80	84	88	91	95	95	96	96	97
12	91	93	94	96	98	98	98	98	98
11	96	97	97	98	99	99	99	99	99
Nr. de linii	18	19	20	21	22	23	24	25	26

Tabelul 15.7

Apeluri/h	Timp mediu de servire, min	Simulare		3+5	
		Linii	Angajaţi	Linii	Angajaţi
100	3,0	7	6	Nu se poate aplica la acest volum	
250	2,0	11	10	12	10
400	2,0	16	14	16	14
250	4,0	19	17	19	17
400	3,0	22	20	22	20
500	3,0	26	24	26	24

pierde după n încercări succesive nereuşite. Estimarea cererii adevărate D, a fost găsită astfel :

$$D = \frac{\text{apeluri acceptate}}{1 - r^n}, \tag{15.1}$$

unde

$$r = \text{rata respingerii} = \frac{C_r}{C_a + C_r} = \frac{\text{apleuri respinse}}{\text{total apeluri}}.$$

În practică, foarte rar se întîmplă ca r să depăşească 15%, iar n poate fi considerat cel puţin egal cu 2. În aceste condiţii, cererea adevărată nu va depăşi niciodată numărul de apeluri acceptate mai

mult de 2%. De aceea, conducătorii au fost sfătuiţi să folosească înregistrările anterioare privind apelurile acceptate pentru a previziona cererea viitoare şi, pe această bază, necesarul de personal.

La acest punct, colegii noştri din departamentul de analiză a încasărilor au preluat rezultatele şi le-au transformat în tabele, ce permit conducătorului să stabilească necesarul de angajaţi în fiecare oră pe timp de o lună, odată ce a fost stabilit numărul de apeluri pe lună (aceasta este posibil datorită stabilităţii variaţiei orare a volumului apelurilor). Rezultatele au fost testate şi au trecut proba cu succes, cu o singură excepţie şi aceasta datorită folosirii incorecte a regulii (s-a folosit „3 plus 5", în loc de „4 plus 7"). Rezultatele au fost implementate* în cadrul sistemului. Partea rămasă nerezolvată din proiect era ordonanţarea orară a angajaţilor, astfel încît să acopere necesităţile oră de oră, să se satisfacă contractele colective şi să fie cît mai economică posibil. Aceasta nu este o aplicaţie a teoriei aşteptării, cu toate că este însăşi rezultatul final al proiectului.

Realitatea este că în general nu este posibil să organizăm ordonanţarea angajaţilor în schimburi în mod automat. Astfel, acurateţea modelului de aşteptare, sau mai curînd a regulilor deduse din el, nu este o chestiune de importanţă deosebită. Cele două reguli operaţionale prezentate — „3 plus 5" şi „4 plus 7" — sînt suficient de corecte pentru a putea fi aplicate. Chiar dacă ele nu ar fi fost atît de corecte, dacă, să spunem ar fi dus la erori mici în 10% din cazuri, adică $\pm$ 1 angajat din 8, nu ar fi trebuit să fim descumpăniţi. Ar fi poate interesant să încheiem capitolul, prezentînd un caz al problemei organizării schimburilor, cu a cărei soluţionare, practic, întregul proiect se încheie.

15.3. ORGANIZAREA SCHIMBURILOR

Odată obţinute necesităţile de personal pentru fiecare secţie sau departament, acestea trebuiau utilizate pentru a selecta modul de organizare a schimburilor şi a determina numărul de angajaţi în fiecare schimb, pentru a asigura acoperirea fiecărui interval individual de timp. Pentru conducător se pun, din nou, două probleme. Prima : organizarea schimburilor trebuie să fie conformă cu politica companiei

* În iunie 1964.

şi cu respectarea duratei schimbului, a numărului de schimburi zilnic, a gradului de utilizare a personalului, a pauzelor de masă etc., aşa cum este specificat în contractele colective de muncă. În al doilea rînd, el trebuie să menţină zilnic numărul minim de angajaţi necesari.

Această problemă este una de programare liniară în numere întregi, în care funcţia obiectiv ce trebuie minimizată este suma numărului de angajaţi de fiecare schimb. Restricţiile conţin condiţiile ca toate apelurile telefonice să poată fi preluate, iar schimburile să fie de durată standard, cu o pauză de masă. Deşi problema poate fi formulată astfel, ea are dimensiuni mari şi nu există metode eficiente, satisfăcătoare.

O aproximare euristică, ce a fost realizată pentru soluţionarea acestei probleme, a dat rezultate excelente şi are avantajul că utilizatorul poate înţelege cît se poate de clar modul de obţinere a soluţiei. Minimul zilnic de angajaţi necesari este dat de un tabel. Dacă se va folosi modul standard de organizare a schimburilor (trei schimburi pe zi : dimineaţa, după amiază şi noaptea), conducătorul va trebui doar să distribuie personal suficient fiecărui schimb, pentru a acoperi cerinţele şi a permite pauzele de masă. Dacă modul de organizare a schimburilor nu este indicat, schimburile pot fi organizate în urma unei perioade de încercări. După o scurtă perioadă de încercări conducătorul va putea obţine rezultate satisfăcătoare, menţinînd un minim de personal în fiecare zi şi utilizînd un număr cît mai mare de schimburi de acelaşi tip.

După ce procedura a fost repetată pentru fiecare zi din săptămînă rezultatele sînt prezentate într-un tabel de alocare în schimburi, care arată numărul de angajaţi necesar pentru fiecare schimb, pe timpul unei săptămîni.

Tabelul 15.8

Tabelul de alocare a personalului în schimburi (exemplu)

Schimbul	Zilele săptămînii							
	D	L	M	M	J	V	S	Total
$00^{01}-08^{00}$ (noaptea)	1	1	1	1	1	1	1	7
$08^{00}-16^{00}$ (zi)	2	2	2	2	2	2	2	14
16^{00} 23^{59} (după amiaza)	2	2	2	2	2	2	2	14
Total	5	5	5	5	5	5	5	35

15.4. REPARTIZAREA PERSONALULUI

Problema ce trebuie rezolvată este de a distribui disponibilul de forță de muncă diferitelor schimburi, astfel încît toate condițiile impuse de tabelul de alocare a personalului în schimburi să fie satisfăcute. De exemplu, tabelul 15.8 poate conduce la tabelul 15.9 (L înseamnă „zi liberă", D „schimbul de dimineață", DA — „schimbul de după amiază", N — „schimbul de noapte").

Tabelul 15.9

O repartizare dedusă din tabelul 15.8

Angajatul	Zilele săptămînii						
	D	L	M	M	J	V	S
A	L	L	D	DA	DA	N	N
B	N	L	L	D	DA	DA	N
C	N	N	L	L	D	DA	DA
D	DA	N	N	L	L	D	DA
E	DA	DA	N	N	L	L	D
F	D	DA	DA	N	N	L	L
G	L	D	DA	DA	N	N	L

Avantajul unei astfel de repartizări este acela că atît conducerea cît și personalul poate ști ușor care persoane vor lucra în timpul unui anumit schimb.

Din păcate, sarcina întocmirii unor astfel de repartizări este îngreunată de un număr de restricții ce pot sau nu să apară la nivel individual. Astfel de restricții sînt:

1. Repartizarea trebuie să se poată roti, adică sarcinile fiecărei persoane trebuie să se poată repeta în cicluri, în cît mai puține săptămîni posibile. Repartizarea dată se rotește cu un ciclu de șapte săptămîni.

2. Orice persoană trebuie să poată lucra în orice ciclu al schimburilor. Astfel, nimănui nu i se permite un tratament preferențial.

3. Dacă este posibil, secvențele succesive de schimburi trebuie să fie identice (o secvență de schimburi este totalitatea acelor schimburi de lucru ce apar între două perioade de „zi liberă").

4. În mod normal, nu se poate întîmpla ca mai mult de 5 sau 6 zile de lucru să separe zilele libere.

5. Zilele libere nu apar singulare, ci în perechi.

6. Schimburi „duble" (adică două schimburi separate de mai puţin de 10 ore) trebuie evitate.

7. Personalul poate avea preferinţa fie pentru secvenţe lungi de schimburi identice (adică de exemplu şase schimburi de noapte succesive), fie pentru secvenţe împărţite în două sau trei tipuri de schimburi (adică două de zi, două de după amiază, două de noapte).

Păstrînd personalul minim într-o secţie, este practic imposibil de construit o repartizare care să ţină seama de toate condiţiile menţionate. Mai mult, rar se întîmplă că fiecare persoană să lucreze numărul corect de ore în fiecare săptămînă. De obicei, orele lucrate săptămînal tind doar în medie să egaleze numărul corect.

Avînd în vedere aceste consideraţii, a fost creat un sistem de reguli pentru întocmirea unor repartizări, pentru a determina soluţia problemelor pentru două oficii de rezervări. Primul pas constă în a calcula numărul minim de angajaţi necesar pentru toată săptămîna. Acesta se obţine prin împărţirea numărului total de schimburi-om necesare (obţinut din tabele de alocare a personalului în schimburi) la 6 pentru o săptămînă de 5 zile lucrătoare.

Rezultatul rotunjit la valoarea întreagă cea mai apropiată se numeşte numărul personalului de bază (NPB). Fiecare schimb-om, determinat de NPB în plus faţă de necesarul total se numeşte schimb fals. În exemplul nostru anterior necesarul total era de 35 de schimburi -om ; NPB a fost egal cu şapte, fără nici un schimb fals.

Pentru a vedea dacă se poate evita apariţia schimburilor false, fiecare necesar zilnic total se scade din NPB şi se obţine numărul de persoane ce vor fi de serviciu în fiecare zi. Un calcul simplu va determina dacă este necesar personal suplimentar peste NPB pentru a evita apariţia unor zile libere singure. Acest calcul va decide de asemenea modul de grupare în perechi a zilelor libere individuale. Rezultatul este un tabel de zile libere cuplate.

Tabelul 15.10 prezintă cuplările ce rezultă din exemplul anterior. Rămîne acum să completăm lista astfel ca secvenţele de zile libere şi de schimburi să fie răspîndite cît mai regulat. Ideal cinci zile lucrătoare trebuie să separe perioade succesive de zile libere, dar ocazional va fi necesar de a avea secvenţe de şase sau chiar şapte zile lucrătoare, pentru a asigura rotaţia repartizărilor. Vom prezenta în continuare cîteva reguli ce au fost stabilite pentru a servi ca ghid conducătorilor ce utilizează metoda :

1. Mai întîi trebuie stabilite cuplările zilelor libere. Acestea se trec pe diagonală în tabelul de repartizare, începînd cu duminica şi lunea primei săptămîni şi terminînd cu sîmbăta şi duminica celei de-a şaptea

Tabelul 15.10

Tabelul numărului de cuplări de zile libere

D/L	L/M	M/M	M/J	J/V	V/S	S/D
1	1	1	1	1	1	1

săptămîni. Se repetă acest mod de repartizare pînă cînd se depăşeşte cel mai mic număr ce apare în tabelul cuplărilor zilelor libere. Apoi se alocă următoarele cuplări astfel încît să se evite secvenţele lungi de schimburi consecutive.

2. Se alocă mai întîi schimburile de noapte şi se continuă cu cele de zi, apoi cu cele de după-amiază.

3. Un schimb de noapte va urma unei zile libere sau unui alt schimb de noapte.

4. Schimburile de zi vor urma după schimburi ce sînt localizate în timp mai devreme în orice secvenţă particulară de schimburi (de menţionat că un schimb de noapte este cel mai timpuriu schimb ce poate apărea în cadrul unei zile).

5. După distribuirea tuturor schimburilor, trebuie să existe numărul corect din fiecare tip de schimb în fiecare zi a spătămînii (aşa după cum se vede în tabelul de alocare a personalului în schimburi).

Concediile de odihnă, precum şi cele de boală sînt considerate în plus faţă de necesarul de bază al personalului şi depind de frecvenţa de îmbolnăvire şi de numărul personalului pentru fiecare situaţie în parte. Numărul personalului suplimentar necesar poate fi uşor obţinut, cunoscîndu-se aceste caracteristici. Numărul acestui personal suplimentar este la latitudinea conducătorului local.

15.5. COMENTARIU

Comentariul acestui capitol este de fapt capitolul 8. Morala sa este că utilitatea unui model de aşteptare este limitată de precizia intrărilor sale. Cu alte cuvinte, ar fi mai util să dedicăm mai mult timp

în a previziona intrările cu suficientă precizie decît a pierde timpul analizînd dacă repartiţia timpului de servire este sau nu exponenţială.

Din cele trei obiective specificate la începutul proiectului, două au fost atinse. Ca rezumat :

1. S-a determinat o formulă pentru a estima necesarul liniilor telefonice şi al angajaţilor în funcţie de previzionări ale volumului de apeluri la oficiile de rezervare de locuri, pentru standarde specificate de servire. Aceasta a fost convertită de către Kirkman şi Mackay de la departamentul de analiză a încasărilor în tabele uşor aplicabile, şi aceste tabele au fost validate de teste făcute de Kirkman. În acest fel, primul obiectiv a fost realizat.

2. S-a dezvoltat, testat şi validat o tehnică euristică pentru planificarea şi organizarea orarului angajaţilor din cadrul oficiilor de rezervare. Astfel, s-a atins al doilea punct.

3. Cea mai eficientă organizare a oficiilor de rezervare nu a fost pînă în prezent suficient cercetată şi determinată.

În viziunea comentariilor făcute în cap. 6 despre eficienţa mult mai mare a simulării cu metoda Monte Carlo, în problemele de comparare a diferitelor forme de organizare a serviciului, decît în acelea privind estimarea numărului de staţii de servire, apare oarecum nefiresc să spunem că al treilea obiectiv continuă să fie un obiectiv. Însă aceasta este situaţia.

16

CONCLUZII

Acum este momentul să ne reîntoarcem la problemele domnilor Szymanowski, Mcdonald şi Overlander, căci sîntem în măsură să examinăm cele trei povestitoare în care şi-au făcut ei apariţia şi să le înţelegem morala. Este clar că inginerul-şef adjunct al corporaţiei „Allied Coke and Iron" nu este prea bine informat despre teoria aşteptării. Căci un sistem de aşteptare în care dublarea capacităţii de servire ar duce doar la îmbunătăţirea timpului mediu de aşteptare este foarte neobişnuit. De exemplu, să considerăm cele două modele $M/M/1 : (\infty/\text{FIFO})$ şi $M/M/2 : (\infty/\text{FIFO})$ în care intrările au aceeaşi intensitate λ. Pentru primul avem $\rho_1 = \lambda/\mu$ iar pentru al doilea $\rho_2 = \lambda/(2\mu)$. Formulele lui Erlang ne dau

$$W_q(1) = \frac{\rho_1}{(1-\rho_1)\,\mu} = \frac{\lambda/\mu}{1-(\lambda/\mu)}\,\frac{1}{\mu}\,; \qquad W_q(2) = \frac{[\lambda/(2\mu)^2}{1-[\lambda/(2\mu)]^2}\,\frac{1}{\mu}\,. \tag{16.1}$$

Astfel, raportul $W_q(2)/W_q(1)$ are expresia

$$\frac{1}{4}\left(\frac{\lambda}{\mu}\right)\frac{1-(\lambda/\mu)}{1-(\lambda/2\mu)^2} \tag{16.2}$$

care, pentru orice valori admisibile ale lui λ şi μ, trebuie să fie $< 1/2$.

Se pare că ideea lui Szymanowski este valoroasă. Dar trebuie să stabilim cît costă instalarea şi funcţionarea unui al doilea cîntar-basculă şi să comparăm acest cost cu cel asociat reducerii timpului de aşteptare al camioanelor. De exemplu, să presupunem că ne costă cu totul A \$ pe zi funcţionarea celui de-al doilea cîntar-basculă şi că un minut de întîrziere al unui camion costă B \$. Atunci reducerea costului datorită instalării celui de-al doilea cîntar-basculă poate fi calculată pentru orice volum al traficului. Să presupunem că într-o oră sosesc în medie 20 camioane şi că timpul de blocare pentru o operaţie de cîntărire este de 2,7 min. Atunci $\lambda/\mu = 0{,}9$ şi un calcul rapid arată că timpul mediu de aşteptare pe camion va fi de 24,3 min în cazul unui singur cîntar-basculă şi de 0,76 min în cazul a două cîntare-basculă. Astfel cîştigul potenţial rezultat din introducerea unui al doilea cîntar-basculă va fi de

$$480 \times 23{,}54 \times B \text{ \$/zi}$$

şi aceasta poate fi comparată cu A pentru a determina cîştigul net al corporaţiei.

Acesta este unul din modurile de a analiza lucrurile, dacă numărul camioanelor aflate în circuit este mare. Un alt mod este de a considera timpul de rotaţie al fiecărui camion şi de a determina sporul de utilizare rezultat din instalarea unui al doilea cîntar-basculă. Există mai multe posibilităţi. Dacă există mai multe puncte de oprire în fiecare circuit al unui camion (şi trebuie să fie cel puţin trei), atunci s-ar putea să fie convenabil modelul ciclic de aşteptare al lui Koenigsberg.

Pe de altă parte, nu este deloc imposibil ca, în ciuda nepriceperii sale în teoria aşteptării, inginerul şef adjunct să posede o altă soluţie mai bună decît instalarea celui de-al doilea cîntar-basculă. Să presupunem că timpul de blocare la cîntărire al unui camion este independent de mărimea camionului (ceea ce este evident pînă la o anumită limită) şi de încărcătură pe care o transportă. Atunci, dacă este posibil să înlocuim camioanele existente cu altele ce pot duce o încărcătură dublă, este posibil să reducem traficul prin această înlocuire. Acelaşi volum de deşeuri poate trece peste cîntarul-basculă de două ori mai repede, adică intensitatea traficului, în erlangi, va fi jumătate din cea veche. Timpul mediu de aşteptare al unui camion se reduce la 2,21 min, iar cîştigul potenţial va depinde de noul cost

al minutului de întîrziere pentru camioanele de acest nou tip ; fie acest cost C $. Atunci cîştigul brut va fi

$$480 \times 24{,}3 \times B - 240 \times 2{,}21 \times C.$$

Costul noilor camioane este, să zicem, D $ (iniţial am presupus că este 250 000 $), astfel că putem calcula uşor cîştigul net. Ar putea fi mai economic să procedăm aşa cum sugerează inginerul-şef adjunct, dar în nici un caz pentru motivele pe care le expune.

Întorcîndu-ne la staţia de cale ferată Paltries, găsim că în principiu propunerea domnului Murdo Mcdonald este o chestiune de priorităţi. În mod formal putem compara rezultatele ce vor fi obţinute dintr-un sistem cu două staţii de servire cu acces nerestricţionat şi două clase de priorităţi relative C_1 şi C_2, cu cele ce vor fi obţinute dintr-un sistem cu două staţii de servire separată, una care serveşte clienţii clasei C_1 şi alta ce serveşte clienţii clasei C_2. Nu este greu să facem calculele teoretic. Dar conducerea staţiei, dacă ar încerca să implementeze sistemul propus, n-ar reuşi niciodată să-l facă să funcţioneze ; căci, să nu uităm, clienţii sînt oameni şi este total improbabil ca ei să lase să intre în faţa lor pe alţii, chiar dacă aceştia ar solicita bilete de prioritate mai mare. Ar fi o rebeliune, sau o obstrucţie împotriva clienţilor din clasa prioritară C_1 care şi-ar cere privilegiile la care au dreptul. Experienţele relatate în cap. 10 ne întăresc în convingerea noastră. Naturam furca expellas, tamen usque recurret*.

Deoarece numărul pasagerilor de clasa întîi este mic în gara Paltries, putem considera evident şi situaţia ca orice pasager din orice clasă să-şi cumpere biletul la oricare din cele două ghişee. Foarte puţini oameni ar suferi din această cauză, fiind lezaţi în superioritatea lor economică, dar multe persoane (printre care şi dl. Mcdonald) ar beneficia de un spor de calitate al serviciului.

Motivele posibile pentru performanţele deplorabile ale serviciului de rezervare de locuri ale companiei aeriene „Transcedental", comparativ cu situaţia din cadrul companiei „Continental and Caribbean" pot fi multe. Cititorul este în măsură acum, după parcurgerea capitolelor 9 şi 15, să completeze o listă lungă de motive ; fără însă a dispune de mai multe informaţii decît a reuşit să ne dea la supărare Overlander, este greu să ajungem la o concluzie limpede. Putem însă să spunem că diferenţa între performanţele celor două servicii are o legătură cu formele de organizare ale acestor servicii. Cititorul este

* Vers din Horaţiu. (*N.T.*)

invitat să privească mai multe organizări posibile, sugerate în fig 16.1. Să presupunem că, oricare ar fi modul de organizare, frecvența apelurilor în orele de vîrf este aceeași, să zicem 30 apeluri pe minut, dar îndrumarea lor este alta.

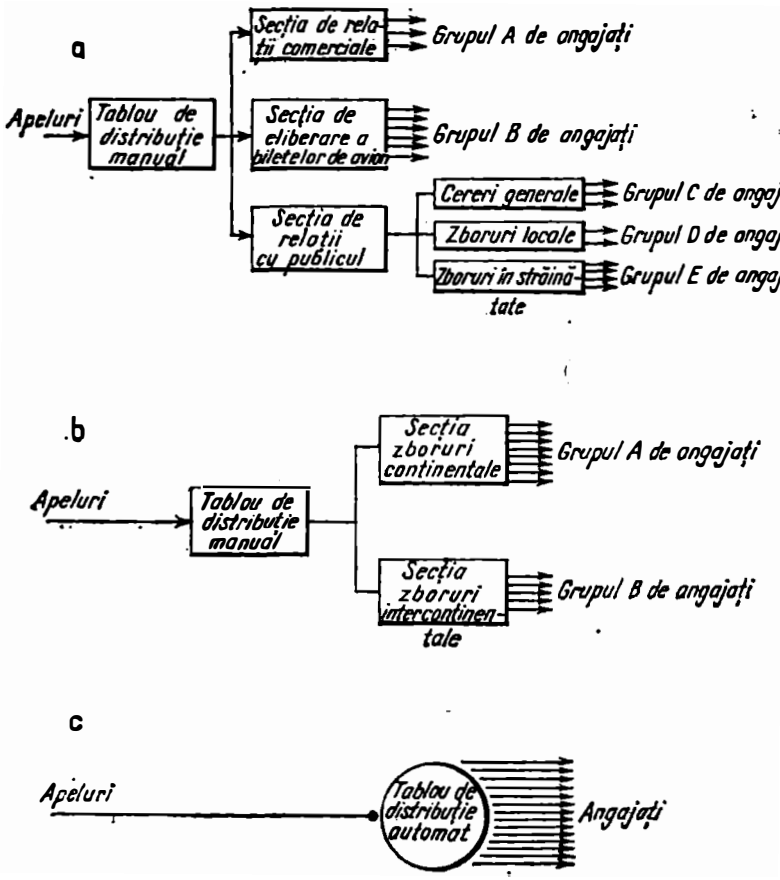


Fig. 16.1. Exemple de sisteme de rezervări telefonice.

Să mai presupunem că repartițiile timpilor de servire sînt toate exponențiale și identice (să zicem cu media 2,0 min). Să ne imaginăm

acum variaţii mari ale numărului de angajaţi în serviciile de rezervare de bilete, ce asigură acelaşi standard de servire, presupunînd peste tot că lungimea şirurilor de aşteptare poate fi infinită; dacă cititorul va voi să adauge şi limitări asupra capacităţii şirurilor de aşteptare, astfel încît va trebui pus în discuţie şi numărul de linii telefonice necesare, nu numai numărul de angajaţi, atunci calculele în scopul comparării vor deveni enorme. Nu se poate obţine uşor un răspuns la întrebarea „De ce compania Continental şi Carbbean asigură un serviciu mai bun decît compania Transcedental? '

Vedem astfel că răspunsurile la trei întrebări legate de trei sisteme de aşteptare, toate aparent simple, diferă substanţial în complexitate şi în volumul de studiu şi calcule pe care-l necesită. Dar din ele — şi să sperăm şi din restul acestei cărţi — putem trage cîteva concluzii utile. Întîi de toate, că însăşi sistemele de aşteptare obişnuite, întîlnite zi de zi, sînt necunoscute pentru majoritatea oamenilor — fie ei ingineri şefi adjuncţi, şefi de gară, diriginţi de oficii poştale, directori de companii aeriene, căpitani de port etc. Cu toată experienţa lor vastă ca clienţi, puţini oameni înţeleg suficient sistemele de aşteptare pentru a putea organiza bine servirea. Experienţa nu ajută la nimic fără pătrunderea fenomenului.

A doua concluzie importantă este că putem face multe pentru îmbunătăţirea servirii cu ajutorul rezultatelor obţinute de teoria aşteptării şi a cîtorva tabele şi grafice. Putem în plus să obţinem soluţii operaţionale folosind modele care, strict vorbind, nu sînt tocmai precise. De asemenea, în mod obişnuit sînt acceptabile modelele aproximative, însă avînd o înţelegere clară şi precisă asupra a ceea ce aproximăm şi cît aproximăm:

Este bine dacă dispunem de modele riguroase în locul unora aproximative; dar este mai bine să avem modele aproximative decît nimic. Este mai bine să avem o teorie logică şi bine fundamentată, însă incompletă şi nedezvoltată, decît să n-avem nici o teorie. Teoria aşteptării, în zilele noastre, este incompletă, dar poate fi folosită, atîta timp cît ne dăm seama de limitele sale. Ea este o teorie fără prea multe baze empirice, bazată în special pe experienţa din domeniul telecomunicaţiilor. Ar fi de dorit să ştim mult mai multe despre comportarea reală a oamenilor în aşteptare; pînă atunci, mare parte din teoria ce apare în revistele de specialitate nu este altceva decît o colecţie de speculaţii matematice fermecătoare. Dar matematica, oricît de ingenioasă ar fi, nu poate ţine locul cunoaşterii umane.

ANEXA 1

LISTA PRINCIPALELOR SIMBOLURI

Convenţiile matematice uzuale sînt menţinute şi în cadrul acestei cărţi; de exemplu, dacă x este o variabilă aleatoare continuă, atunci $f(x)$ este densitatea sa de repartiţie, iar $F(x)$ este funcţia de repartiţie.

Dacă însă y este o variabilă deterministă, atunci $g(y)$ este o funcţie de y.

Transformatele Laplace sînt notate prin asteriscuri. Astfel, transformata Laplace a funcţiei $F(x)$ se scrie $F^*(z)$.

În cadrul textului au fost utilizate anumite simboluri speciale, folosite doar cu un singur sens în întreaga carte: acestea formează lista de mai jos. Celelalte simboluri sînt utilizate cu diferite semnificaţii, evident cu precizarea sensului pe care îl au în fiecare context.

a, $a(t)$, $A(t)$	media, densitatea de repartiţie şi funcţia de repartiţie a timpilor dintre sosirile clienţilor;
b, $b(t)$, $B(t)$	media, densitatea de repartiţie şi funcţia de repartiţie a timpilor de servire (rau de reţinere);
c	numărul staţiilor de servire;
D	se referă la o repartiţie constantă;
E_k	se referă la repartiţia Erlang de parametrul k, adică repartiţia sumei a k variabile independente şi repartizate exponenţial identic;
G	se referă la o repartiţie generală;
GI	se referă la o repartiţie generală cu ipoteze de independenţă;

L	capacitate a sistemului, adică limita superioară a numărului de clienţi ce pot exista în sistem în orice moment;
n, $\bar{n}$	numărul clienţilor existenţi în sistem şi valoarea sa medie;
n_q, $\bar{n}_q$	numărul clienţilor existenţi în sistem şi care aşteaptă să fie serviţi, respectiv media lor;
$p_n(t)$	probabilitatea existenţei a n clienţi în sistem, la momentul t;
$P(s, t)$	funcţia generatoare a probabilităţilor $p_n(t)$;
p_n	probabilitatea, în starea stabilă, a existenţei a n clienţi în sistem;
$P(s)$	funcţia generatoare a probabilităţilor p_n;
s	o variabilă fictivă ce apare în cadrul funcţiilor generatoare;
u	variabilă statistică, ce măsoară timpul total de aşteptare;
u	utilizarea staţiilor de servire;
v	variabilă statistică măsurînd timpul de aşteptare;
V	coeficientul de variaţie a unei repartiţii;
W, $w(u)$, $W(u)$	media, densitatea de repartiţie şi funcţia de repartiţie a timpilor totali de aşteptare a clienţilor în sistem;
W_q, $w_q(v)$, $W_q(v)$	media, densitatea de repartiţie şi funcţia de repartiţie a timpilor de aşteptare a clienţilor în sistem;
z	variabila unei transformate Laplace;
δ_{ij}	simboluri lui Kronecker;
λ	intensitatea medie a sosirilor clienţilor;
μ	intensitatea medie de servire (pentru o staţie de servire);
ρ	intensitatea traficului, exprimată în erlangi;
ε	$= \lambda/\mu$ traficul oferit, exprimat în erlangi.

ANEXA 2

FORMULE ŞI GRAFICE UTILE

Această anexă este împărţită în două părţi. În prima parte sînt prezentate anumite grafice legate de modelele de aşteptare întîlnite în cadrul cărţii. În partea a doua se prezintă tabele ale principalelor formule întîlnite în cadrul modelelor cel mai des utilizate.

GRAFICE ŞI CURBE DE LUCRU

Există şapte grafice: primele patru, de la A 2-1 la A 2-4, sînt reproduse după lucrarea:

Burke, P. J. *Equilibrium Delay Distribution for One-Channel with Constant Holding Time, Poisson Input and Random Service.* În: Bell System Technical Journal, July 1959.

Aduc şi pe această cale mulţumirile mele editorilor revistei „Bell System Technical Journal" pentru permisiunea publicării acestor reproduceri. Aceste grafice au fost realizate de Burke pentru a prezenta rezultatele pentru modelul $M/D/1$: (∞/SIRO), el fiind primul care l-a studiat. Un interes suplimentar îl prezintă faptul că în graficele trei şi patru sînt încorporate şi rezultatele referitoare la modele $M/D/1$: (∞/FIFO) şi respectiv $M/M/1$: (∞/SIRO). În aceste grafice, ca şi în cele ce urmează, timpii de aşteptare (sau întîrzierile) sînt exprimaţi ca multipli ai timpului mediu de servire (sau de reţinere).

Aceasta este o practică frecventă în graficele de aşteptare. În grafice se utilizează de asemenea termenul de grad de ocupare care în cazul modelelor ce permit şiruri de aşteptare infinite, fără respingeri de clienţi, este identic cu termenul de intensitate a traficului.

Graficele A2-5 — A2-7 sînt reproduse cu permisiunea editorilor după lucrarea :

Wilkinson, R. I. *Working Curves for Delayed Exponential Calls Served in Random Order.* În : Bell System Technical Journal, March 1953.

Şi aceste grafice au un domeniu mai larg de aplicabilitate decît sugerează titlul lucrării. Primele două pot fi aplicate la orice model din familia $M/M/c$:(∞/GD), unde GD semnifică că orice fel de disciplină de aşteptare, cu excepţia celor rezultate din respingerea clienţilor. Ele sînt aplicabile în special principalelor trei modele cu disciplina de aşteptare FIFO, SIRO şi LIFO. Ultimul grafic, A3-7, pe care Wilkinson l-a preluat dintr-o lucrare a lui Riordan, anume „Delay Curves for Calls Served at Random". În : Bell System Technical Journal, January 1953, se referă doar la modelul $M/M/c$: (∞/SIRO).

Lucrarea lui Wilkinson conţine şi alte grafice foarte utile, referitoare la modelele specifice de acest tip, cum ar fi pentru $c = 1, 2, \ldots$, pe care însă nu le reproducem aici. Nu sînt necesare explicaţii suplimentare referitoare la graficele prezentate. Ultimul ar putea fi însă o excepţie. Pentru a obţine probabilitatea ca fiecare client să aştepte mai mult decît v unităţi de timp, trebuie calculat v/b, care este timpul aşteptării exprimat ca multiplu al timpului mediu de servire, iar apoi trebuie calculat $R = cv/b$. Atunci

$1 - W_q(v)$ = probabilitatea unui timp de aşteptare mai lung decît $v = W_q(>0) \cdot Q(>R)$,

unde $W_q(>0)$ şi $Q(>R)$ sînt funcţiile prezentate în graficile A3-6 şi respectiv A3-7.

FORMULE UTILE

Această parte este destinată exclusiv prezentării unor formule foarte utilizate în teoria aşteptării. Pentru anumite modele există formule generale, dar ele nu sînt prezentate aici, deoarece sînt dificil de înţeles şi de calculat.

1. *O staţie de servire, sosiri repartizate Poisson* $M/G/1$: (∞/GD).

Formulele lui Pollaczek — Hincin pot fi utilizate pentru estimarea numărului mediu de clienţi existenţi în sistem şi a timpului

mediu de aşteptare. Alte elemente nu se obţin cu uşurinţă, exceptînd cazurile particulare în care repartiţia timpului de servire este exponenţială, constantă sau erlangiană :

$$n = \rho + \frac{\rho^2(1 + V_b^2)}{2(1 - \rho)}, \qquad W_q = \frac{\rho(1 + V_b^2)}{2(1 - \rho)\mu}.$$

$M/M/1 : (\infty/GD)$

Repartiţiile numărului de clienţi în aşteptare (în şirul de aşteptare sau în sistem) nu sînt dependente de disciplina de aşteptare. Soluţia în starea de tranziţie este

$$p(n, t) = \exp\{-(\lambda + \mu)t\} \times$$

$$\times \left[\rho^{n/2} I_n(x) + \rho^{(n-1)/2} I_{n+1}(x) + \rho^n(1-\rho) \sum_{r=2}^{\infty} \rho^{-(n+r)/2} I_{n+r}(x) \right],$$

unde $x = 2t\sqrt{\lambda\mu}$. Soluţia în starea stabilă este

$$p_n = (1 - \rho)\,\rho^n, \qquad n = 0, 1, 2, \ldots$$

Timpul total mediu de aşteptare şi timpul mediu de aşteptare nu sînt afectaţi de natura disciplinei de aşteptare. Ei sînt

$$W_q = \frac{\rho}{1-\rho} \cdot \frac{1}{\mu}, \qquad W = \frac{1}{1-\rho} \cdot \frac{1}{\mu}.$$

Repartiţiile timpilor de aşteptare şi a timpilor totali de aşteptare sînt afectaţi de disciplina de aşteptare Avem :

—FIFO

$$w_q(v) = \begin{cases} 1 - \rho, & v = 0, \\ \rho(1-\rho)\,\mu\, e^{-(1-\rho)\mu v}, & v > 0, \end{cases}$$

$$w(u) = (1 - \rho)\,\mu\, e^{-(1-\rho)\mu u}, \qquad u \geqslant 0.$$

— SIRO

Probabilitatea ca un client care trebuie să aştepte mai mult decît t este dată de aproximarea lui Riordan (v. graficul A3-7) :

$$\frac{\alpha}{2}\exp\{-\alpha(1-\rho)\,\mu\,t\}+\frac{\beta}{2}\exp\{-\beta(1-\rho)\,\mu t\},$$

unde $\alpha = 1+\sqrt{\rho/2}$, iar $\beta = 1-\sqrt{\rho/2}$.

— LIFO

Repartiţia timpilor totali de aşteptare este identică cu repartiţia perioadelor ocupate pentru modelul FIFO. Densitatea de repartiţie este

$$w(u) = \exp\{-(1-\rho)\,\mu\,v\}\,\frac{I_1(2v\sqrt{\lambda\mu})}{v\sqrt{\rho}}\,.$$

$M/E_k/1$: (∞/FIFO).

Repartiţia numărului de clienţi existenţi în sistem este foarte uşor exprimată cu ajutorul funcţiei generatoare ; din aceasta diferitele valori ale lui p_n pot fi obţinute prin derivare repetată :

$$P(s) = \frac{(1-k\rho)(1-s)}{1-s-\rho s(1-s)^k}\,.$$

Transformata Laplace a repartiţiei timpilor de aşteptare este dată de

$$P\left(\frac{1}{1+z/\mu}\right).$$

O tratare analitică a modelului este destul de anevoioasă. Exceptînd cazul $k = 1$, calculul inversiei transformatei Laplace necesită anumite funcţii speciale, studiate de Luchak, care se aseamănă într-o oarecare măsură cu funcţiile Bessel incomplete.

Repartiţia perioadei ocupate ridică aceleaşi probleme şi de aceea nu există o formulă utilizabilă pentru modelul $M/E_k/1$: (∞/LIFO).

$M/D/1$:(∞/FIFO)

Probabilităţile în starea stabilă sînt date de

$$p_n = (1-\rho)\sum_{k=1}^{n}(-1)^{n-k}\,e^{k\rho}\left[\frac{(k\rho)^{n-k}}{(n-k)!}+\frac{(k\rho)^{n-k-1}}{(n-k-1)!}\right].$$

Repartiţia timpilor de aşteptare este dată de

$$W_q(v) = (1-\rho)\sum_{i=0}^{k}\exp\left[\rho(\mu v-i)\right]\frac{[\rho(\mu v-i)]^i}{i!}\,.$$

Pentru a obţine mediile, se pot utiliza formulele lui Pollaczek—Hincin. Valorile medii cele mai utilizate sînt

$$\bar{n} = 2 + \frac{1}{2} \cdot \frac{\rho}{1-\rho}, \qquad W_q = \frac{1}{2} \cdot \frac{\rho}{1-\rho} \cdot \frac{1}{\mu}.$$

$M/D/1 : (\infty/\text{SIRO})$

Formulele acestui model nu s-au obţinut încă; vezi graficele A3-1 — A3-4.

SERVICIU PRIORITAR

Singurele formule utilizabile sînt acelea ale mediilor. Pentru două clase de priorităţi relative avem :

$$\bar{n}_1 = \frac{\alpha\rho(1+\rho-\alpha\rho)}{1-\alpha\rho}, \qquad \bar{n}_2 = \frac{(1-\alpha)\,\rho(1-\alpha\rho+\alpha\rho^2)}{(1-\rho)(1-\alpha\rho)},$$

$$\bar{n}_{q1} = \frac{\alpha\rho^2}{1-\alpha\rho}, \qquad \bar{n}_{q2} = \frac{(1-\alpha)\,\rho^2}{(1-\rho)(1-\alpha\rho)},$$

$$W_{q1} = \frac{\lambda}{\mu(\mu-\alpha\lambda)}, \qquad W_{q2} = \frac{\lambda}{(\mu-\lambda)(\mu-\alpha\lambda)}.$$

Pentru două clase de prioritate, din care una are prioritate absolută faţă de cealaltă, avem :

$$p_{n_1} = \left(\frac{\alpha_1\lambda}{\mu_1}\right)^{n_1}\left(1-\frac{\alpha_1\lambda}{\mu_1}\right), \qquad n_1 = 0, 1, 2, \ldots,$$

$$\bar{n}_1 = \frac{\rho_1}{1-\rho_1}, \qquad \bar{n}_2 = \frac{\rho_2}{1-\rho_1-\rho_2}\left(1+\frac{\rho_1}{1-\rho_1}\,\frac{\mu_2}{\mu_1}\right),$$

unde $\rho_1 = \alpha_1\lambda/\mu_1$,

$$W_1=\left(1+\frac{1}{1-\rho_1}\right)\frac{1}{\mu_1}, \qquad W_2=\frac{1}{\mu_2(1-\rho_1-\rho_2)}\left(1+\frac{\rho_1}{1-\rho_1}\cdot\frac{\mu_2}{\mu_1}\right),$$

$$W_1=\left(1+\frac{1}{1-\rho_1}\right)\frac{1}{\mu_1}, \qquad W_2=\frac{1}{\mu_2(1-\rho_1-\rho_2)}\left(1+\frac{\rho_1}{1-\rho_1}\cdot\frac{\mu_2}{\mu_1}\right),$$

$$W_{q1}=\frac{\rho_1}{\mu_1(1-\rho_1)}, \qquad W_{q2}=\frac{1}{\mu_2(1-\rho_1-\rho_2)}\left(\rho_1+\rho_2+\frac{\rho_1}{1-\rho_1}\cdot\frac{\mu_2}{\mu_1}\right).$$

MAI MULTE STAȚII DE SERVIRE, SOSIRI REPARTIZATE POISSON

M/M/c : (∞/FIFO)

Pentru acest model există grafice foarte utile. Graficele A3-5 și A3-6 sînt ambele utilizabile. În plus, există graficele lui Molina, mult mai detaliate, ce pot fi găsite în lucrarea :

Molina, E. C. *Application of the Theory of Probability to Telephone Trunking Problems.* În : Bell System Technical Journal, July, 1927.

De aceea formulele nu sînt de mare folos practicienilor. Principalele formule sînt :

$$p_0 = \frac{1}{\sum_{r=0}^{c-1}\frac{(c\rho)^r}{r!}+\frac{(c\rho)^c}{c!}\cdot\frac{1}{1-\rho}}, \qquad p_1 = \begin{cases} \dfrac{(c\rho)^n}{n!}p_0, & 1 \leqslant n \leqslant c; \\ \dfrac{c^c}{c!}\,\rho^n p_0, & n > c. \end{cases}$$

Probabilitatea ca un client să sosească și să găsească toate stațiile de servire ocupate este

$$\sum_{r=c}^{\infty} p_r = \frac{(c\rho)^c}{c!(1-\rho)}p_0.$$

Această cantitate este frecvent notată în literatura prin simbolul

$P(>0)$. În carte ea este notată $Q(0)$ şi este egală cu complementul lui $W_q(0)$, care este probabilitatea unui timp de aşteptare pozitiv. Similar $Q(v)$ este probabilitatea unui timp de aşteptare ce depăşeşte v, care este complementul lui $W_q(v)$. Această cantitate este notată de obicei prin $P(>t)$. Atunci

$$Q(v) = \exp\ [-c\mu\,(1-\rho)\,v]\cdot Q(0).$$

În plus :

$$\bar{n}_q = \frac{\rho(c\rho)^c}{c!\,(1-\rho)}\,p_0, \qquad W_q = \frac{\bar{n}_q}{\lambda}.$$

$M/M/c$: (L/FIFO)

Nu există grafice utile pentru acest model. Principalele formule sînt

$$p_0 = \frac{1}{\sum_{r=0}^{c-1}\frac{(c\rho)^2}{r!} + \frac{(c\rho)^c}{c!}\left[\frac{1-\rho^{L-c+1}}{1-\rho}\right]}, \qquad p_n = \begin{cases} \dfrac{(c\rho)^n}{n!}, & 1 \leqslant n \leqslant c, \\ \dfrac{c^c}{c!}\,\rho^n p_0, & c < n \leqslant L. \end{cases}$$

Cantitatea p_L este probabilitatea ca un client să fie respins (aşa-numita „probabilitate de pierdere").

$M/M/c$: (∞/SIRO)

Nu merită să prezentăm formula pentru acest model. Formulele exacte sînt nefolosibile în scopuri practice. Graficele A3-5 — A3-7 sînt mult mai adecvate.

$M/D/c$:(∞/FIFO)

Pentru acest model există grafice deosebit de utile privind repartiţia timpului de aşteptare şi a timpului mediu de aşteptare. O sursă bibliografică extrem de interesantă este

Shelton, J. R. *Solution Methods for Waiting Line Problems.* În : Journal of Industrial Engineering, **11**, 4, 1960,

Există prea multe grafice pentru a fi reproduse aici. Ele sînt realizate pentru $c = 1, 2, 3, 4, 5, 8, 10, 20$. Singurele formule care pot fi aplicate sînt următoarele :

$$P(s) = -\frac{c-\lambda}{(1-s_1)(1-s_2)\ldots(1-s_{c-1})}\cdot\frac{(s-1)(s-s_1)\ldots(s-s_{c-1})}{1-s^c\,e^{\lambda(1-s)}},$$

unde $1, s_1, s_2, \ldots, s_{c-1}$ sînt cele c rădăcini ale ecuației

$$1 - s^c e^{\lambda(1-s)} = 0,$$

și pentru timpul mediu de așteptare:

$$W_q = \sum_{i=1}^{\infty} e^{-i\lambda} \left[\sum_{j=ic}^{\infty} \frac{(i\lambda)^j}{j!} - \frac{c}{\lambda} \sum_{j=ic+1}^{\infty} \frac{(i\lambda)^j}{j} \right].$$

STAȚII DE SERVIRE ÎN SERIE

Există puține rezultate aplicabile. Singurele formule cu semnificație practică sînt acelea, datorate lui Jackson, legate de un sistem cu k faze, fiecare de tipul $M/M/c$: (∞/FIFO) :

$$p(n_1, n_2, \ldots, n_k) = p(0, 0, \ldots, 0) \prod_{r=1}^{k} b(n_r),$$

unde n_r este numărul clienților prezenți în faza r, iar

$$b(n_2) = \begin{cases} \dfrac{(c_r \rho_r)^{n_r}}{n_r!}, & n_r < c_r ; \\ \dfrac{c_r^{c_r}}{c_r!} \rho_r^{n_r}, & n_r \geqslant c_r, \end{cases}$$

ρ_r fiind $\lambda/(c_r \mu_r)$.

STAȚII DE SERVIRE CICLICE

Formulele lui Koenigsberg pentru un sistem ciclic cu k faze sînt

$$p(n_1, n_2, \ldots, n_k) = \frac{\lambda_1^{n_1} \lambda_2^{n_2} \ldots \lambda_k^{n_k}}{\sum_\pi \lambda_1^{r_1} \lambda_2^{r_2} \ldots \lambda_k^{r_k}},$$

unde $\lambda_i = 1/\mu_i$ iar Σ_π semnifică însumarea după toate combinațiile, astfel încît $\sum_{i=1}^{k} r_i = N$, unde N este numărul clienților existenți în sistem. Probabilitatea ca faza i să fie liberă este $\dfrac{k-1}{N+k-1}$.

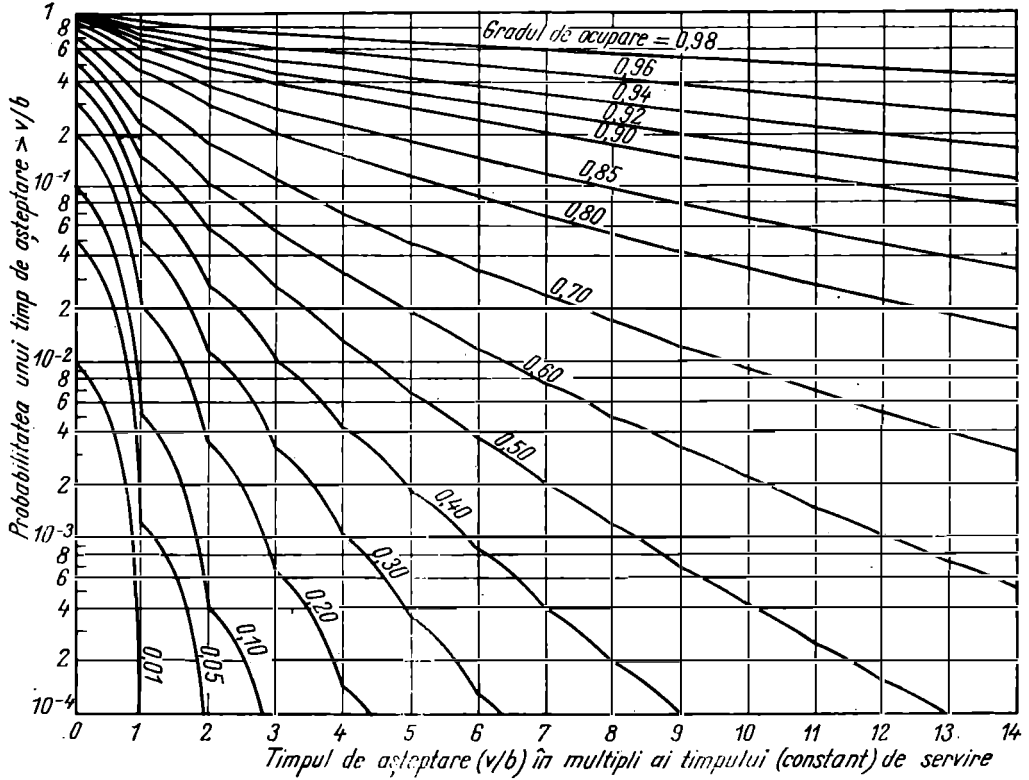


Fig. A2-1. $M/D/1$; (∞ SIRO); repartiția timpului de așteptare.

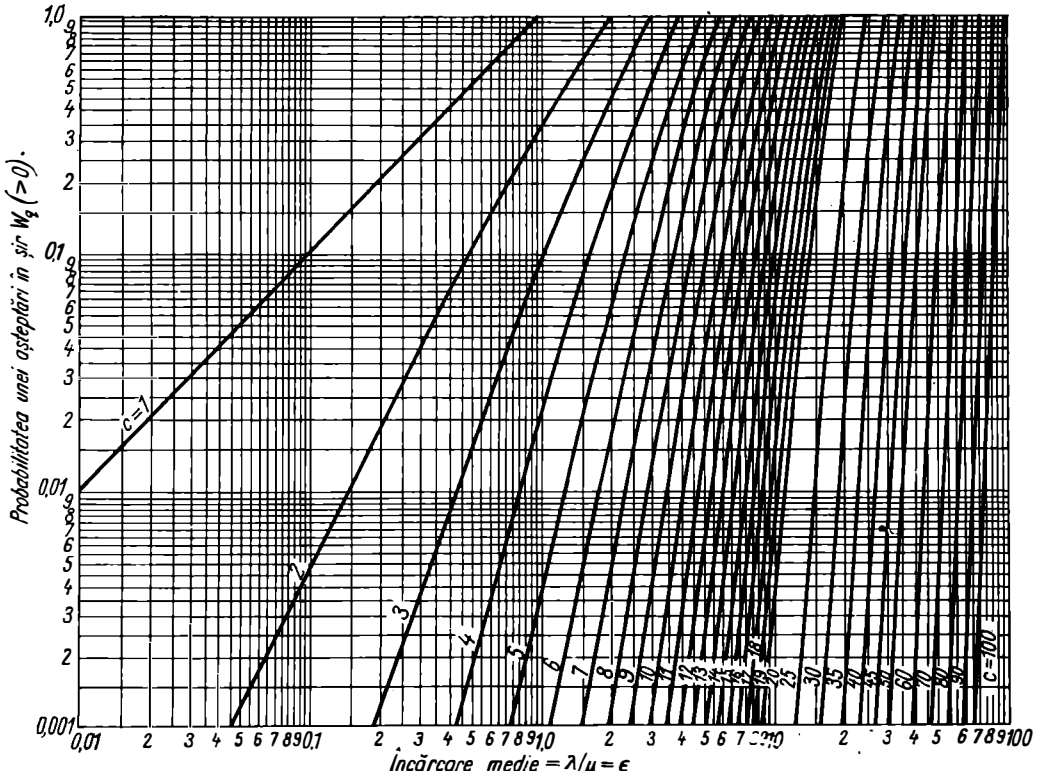


Fig. A2-6. Probabilitatea așteptării (a ocupării simultane a tuturor stațiilor de servire) pentru $M/M/c$: (∞/GD).

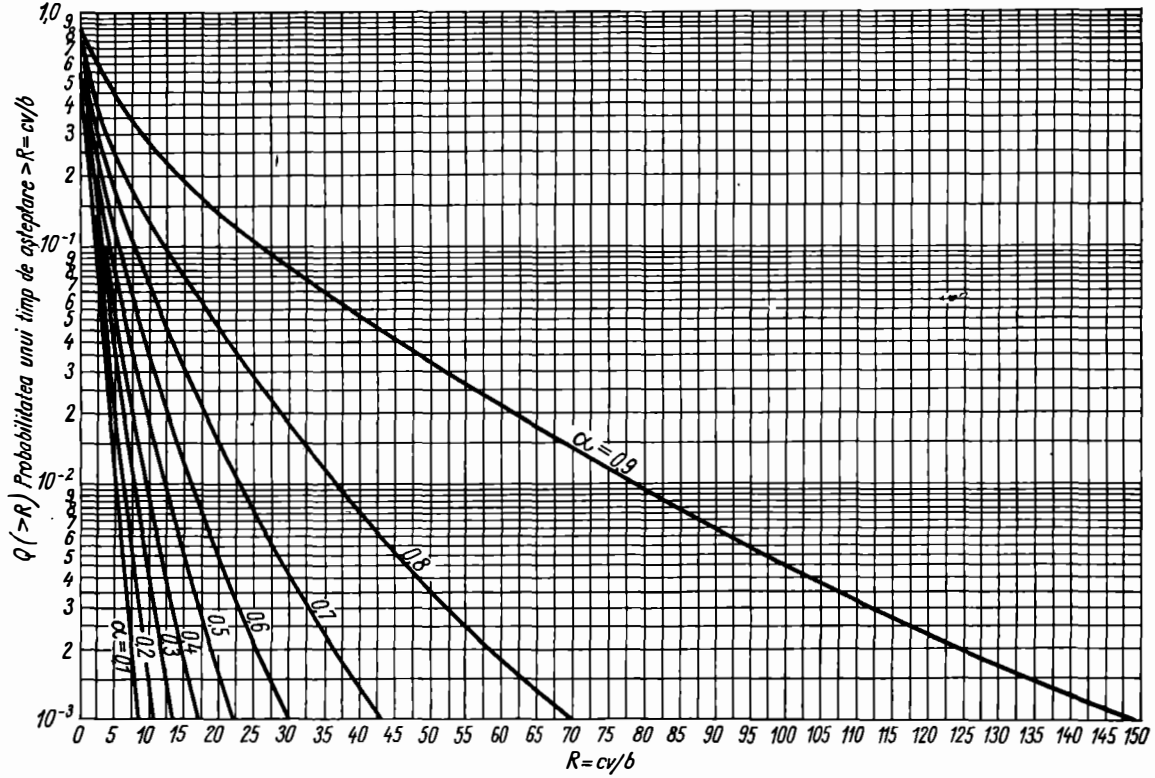


Fig. A2-7. $M/M/c$: $(\infty/SIRO)$: repartiţia timpilor de aşteptare pentru clienţii nevoiţi să aştepte.

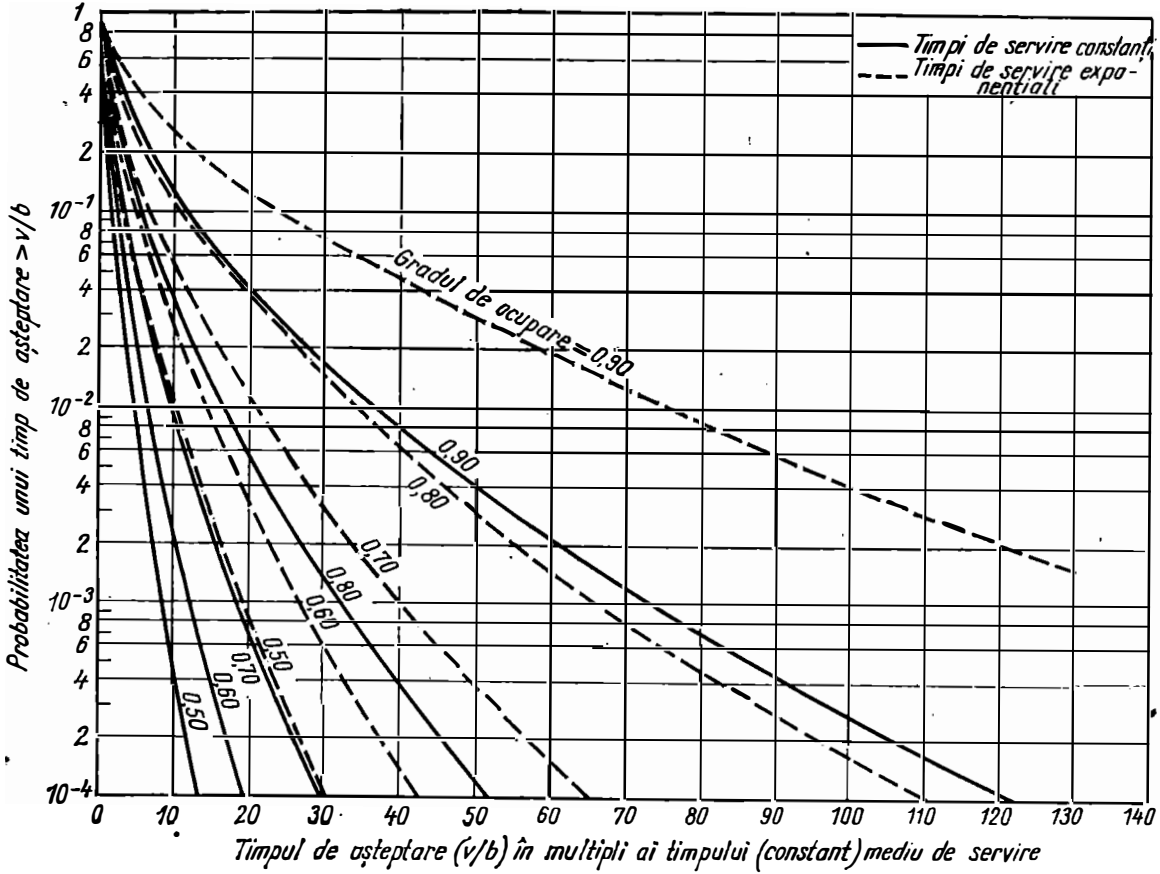


Fig. A2-1. Comparație între repartițiile timpilor de așteptare pentru $M/D/1 : (\infty/\mathrm{SIRO})$ și $M/M/1 : (\infty/\mathrm{SIRO})$.

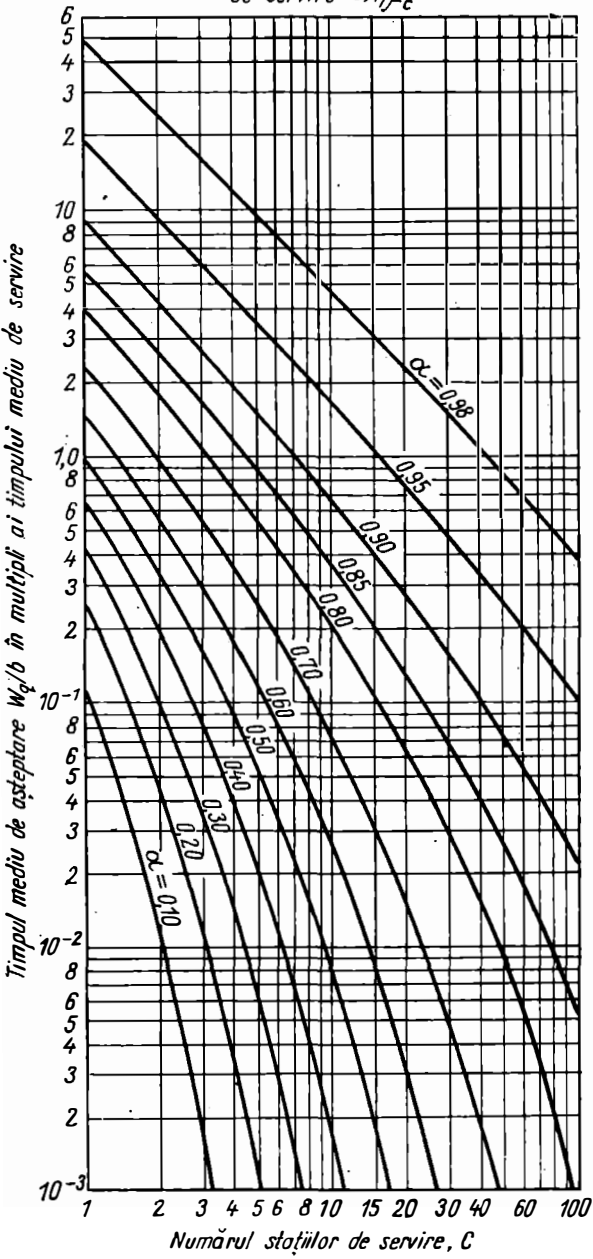


Fig. A2-5. Timpul mediu de așteptare pentru $M/M/c$: (∞/SIRO).

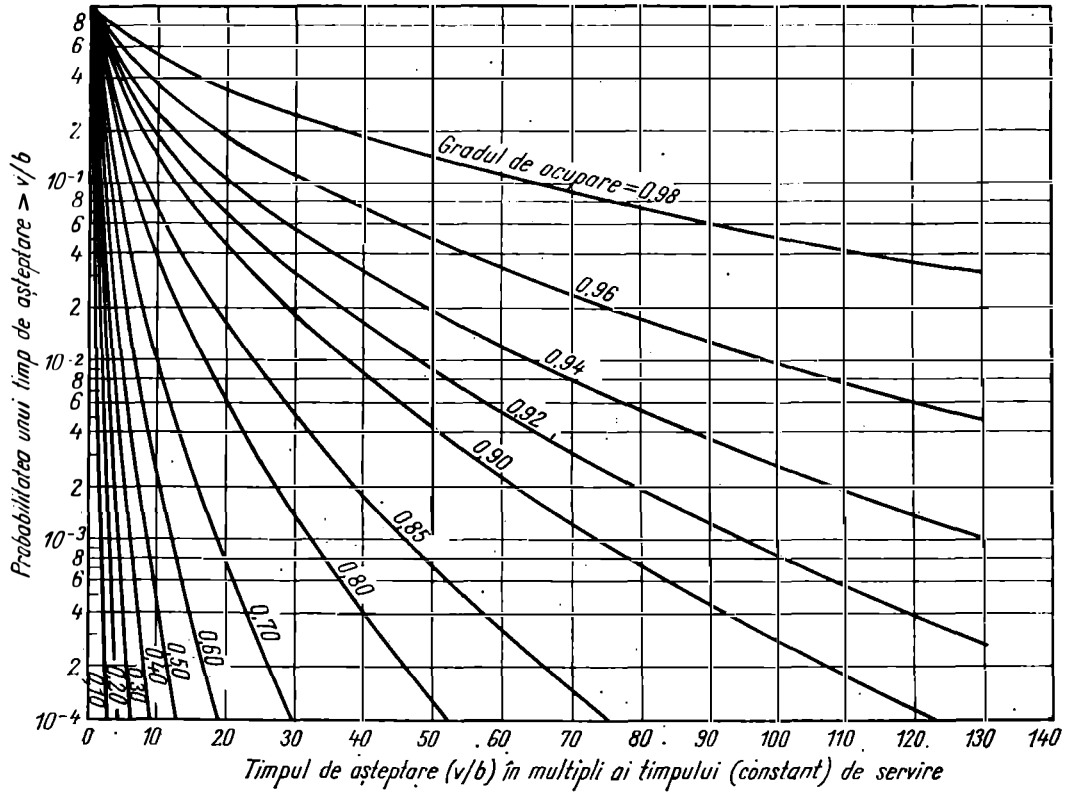


Fig. A2-2. $M/D/:\infty/(\mathrm{FIFO})$: repartiția timpului de așteptare.

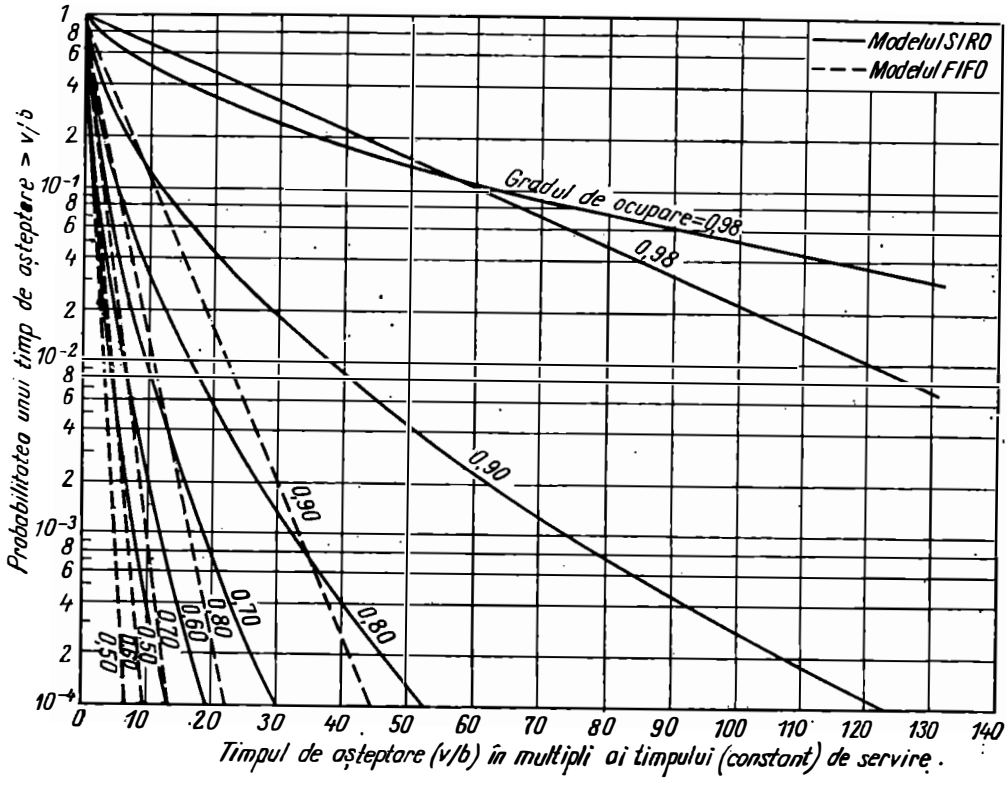


Fig. A2-3. Comparație între repartițiile timpilor de așteptare corespunzătoare modelelor $M/D/1$: (∞/SIRO) și $M/D/1$: (∞/FIFO).

BIBLIOGRAFIE

Lucrări citate de către autor:

1. Cox, D. R., Smith, W. L. *Queues.* London, Methuen, 1960.
2. Saaty. T. L. *Elements of Queueing Theory with Applications.* New York, McGraw-Hill, 1961.
3. Morse, P. *Queues, Inventories and Maintenance.* New York, Wiley, 1958.
4. Feller, W. *An Introduction to Probability Theory and its Applications.* New York, Wiley, 1957.
5. Kendall, M. G., Stuart, A. *The Advanced Theory of Statistics.* I New York, Hafner, 1958.
6. Widder, D. V. *The Laplace Transform.* Princeton, Princeton Univ. Press, 1949.
7. Tocher, K. D. *The Art of Simulation.* London, English Univ. Press, 1963.
8. Oxenfeldt, A. E. (Ed.) *Models of Markets.* New York-London, Columbia Univ. Press, 1963.
9. Edie, L. C. *Trafic Delays at Toll Booths.* În: Op. Research, vol. 2, bȓ. 2, 1954.
10. Friend, J. E. *Two Studies in Airport Congestion.* În: Oper. Res. Quarterly, vol. 9, nr. 3, 1958.
11. Fry, J., Jackson, R. R. P., Welch, J. D. *Appointment Systems in Hospitals and General Practice.* În: Oper. Res. Quarterly, vol. 15, nr. 3, 1964.
12. Gebhart, R F. *A Limiting Distribution of an Estimate of Mean Queue Lenght.* În: Op. Research, vol. 11, nr. 6, 1963.
13. Lee, A. M. *Some Aspects of a Control and Communication Systems.* În: Oper. Res. Quarterly, vol. 10, nr. 4, 1959.
14. Lee, A. M., Linder, R. W. *Two Problems in Scheduling Manpower.* În: J. Canad. Oper, Res. Soc., vol. 3, nr. 1, 1965.
15. Lee, A. M., Longton, P. A. *Queueing Processes Associated with Airline Passenger Check-in.* În: Oper. Res. Quarterly, vol. 10, nr. 1, 1959.
16. Molina, E. D. *Application of the Theory of Probability to Telephone Trunking Problems.* În: Bell System Teleph. J., vol. 6, July, 1925.
17. Shelton, J. R. *Solution Methods for Waiting Line Problems.* În: J. Industrial Eng., vol. 2, nr. 4, 1960.
18. Stillson, P. *Implementation Problems in O.R.* În: Op Research, vol. 11, nr. 1, 1963.
19. Tocher, K. D. *The Rôle of Models in Operational Research.* În: J. Royal Stat. Soc., Part II, 1961.
20. Wilkinson, R. I. *Working Curves for Delayed Exponential Calls Served in Random Order.* În: Bell System Teleph. J., vol. 34, 1960.

Cititorul poate consulta, pentru explicarea noțiunilor teoretice utilizate, următoarele lucrări în limba română:

1. Mihoc, Gh., Ciucu, G. *Introducere în teoria așteptării.* Editura tehnică, București, 1967.
2. Mihoc, Gh., Ciucu, G., Muja, A. *Modele matematice ale așteptării.* București, Editura Academiei R.S.R., 1973.
3. Ciucu, G. *Elemente de teoria probabilităților și statistică matematică.* București, Editura didactică și pedagogică, 1963.
4. Iosifescu, M., Mihoc, Gh. ș.a. *Teoria probabilităților și statistica matematică.* București, Editura tehnică, 1966.
5. Mihoc, Gh., Ciucu, G., Craiu, V. *Teoria probabilităților și statistică matematică.* București, Editura didactică și pedagogică, 1970.
6. Kaufmann, A. *Metode și modele ale cercetării operaționale.* Vol. I, București, Editura științifică, 1967.

CUPRINS

Prefață 5

1. INTRODUCERE 7

1.1. Problema oțelăriilor 8
1.2. Problema casei de bilete 8
1.3. Problema soției furioase 9
1.4. Natura sistemelor de așteptare 10

2. MODELE ALE SISTEMELOR DE AȘTEPTARE 12

2.1. Generalități 12

2.2. Caracteristicile modelelor de așteptare 14
3. MODELE CU O SINGURĂ STAȚIE DE SERVIRE 30

3.1. Generalități 30
3.2. Modele cu timpi de servire exponențiali 31
3.3. Modele cu timpi de servire neesponențiali 43
3.4. Modele cu timpul de servire constant 49
3.5. Reguli de priorități pentru disciplina de așteptare 53
3.6. Comentariu general 58

4. MODELE CU MAI MULTE STAȚII DE SERVIRE 59

4.1. Modele 59
4.2. Fenomene de deplasare 70

5. MODELE ALE SISTEMELOR SERIALE ȘI CICLICE 74

5.1. Sisteme seriale 75
5.2. Sisteme ciclice 82

6. TEHNICI DE SIMULARE 85

6.1. Metode Monte Carlo 85
6.2. Numere aleatoare și pseudoaleatoare 88
6.3. Generarea variațiilor aleatoare 89
6.4. Utilizarea simulărilor 92
6.5. Experimente de simulare 93
6.6. Concluzie 97

7. NATURA SOLUȚIILOR OPERAȚIONALE 98

8. TREI TIPURI DE APROXIMĂRI 106

9. SISTEME DE REZERVARE DE LOCURI 111

9.1. Descrierea problemei 111
9.2. Prima fază 113

9.3. Interludiu 118
9.4. Faza a doua 119
9.5. Comentariu 123

10. METODE DE CONTROL AL CĂLĂTORILOR 124

10.1. Prima fază 125
10.2. Faza a doua 134
10.3. Studiul de validare 147
10.4. Comentariu 152

11. PROBLEMA SORTĂRII CUPOANELOR 153

12. PATRU SITUAȚII SELECTATE 158

12.1. Proiectarea rețelei de piste în aeroport 158
12.2. Problema autobuzelor în aeroport 160
12.3. Întîrzieri ale circulației în punctele de vamă 161
12.4. Sisteme de întîlnire 164

13. PROIECTAREA SEDIULUI DIN LONDRA AL COMPANIEI BEA 166

13.1. Prezentarea problemei 166
13.2. Capacitatea locului de parcare 168
13.3. Împrăștierea pasagerilor ce părăsesc sediul 171
13.4. Încărcarea autobuzelor spre aeroport 180

14. PROBLEME DE AȘTEPTARE ÎNTR-UN SISTEM DE CONTROL 186

14.1. Prezentare a problemei 187
14.2. Studiul preliminar 188
14.3. Studiul principal: controlul plecărilor 190

15. O PROBLEMĂ A STANDARDELOR DE SERVIRE 195

15.1. Prima fază 196
15.2. Faza a doua 211
15.3. Organizarea schimburilor 217
15.4. Repartizarea personalului 219
15.5. Comentariu 221

16. CONCLUZII 223

Anexa 1. LISTA PRINCIPALELOR SIMBOLURI 228

Anexa 2. FORMULE ȘI GRAFICE UTILE 230

Bibliografie 245

Redactor: VALENTINA CREŢU
Tehnoredactor: ELENA GERU
Coperta şi supracoperta seriei: VALENTIN VIŞAN

Bun de tipar: 14.06.1976. Coli de tipar: 15,50. Tiraj: 5350+50 exemplare legate. C.Z.: 517.8:519.9.

c. 182 I. P. INFORMAŢIA
Str. Brezoianu 23–25.
Bucureşti

Lei 18

În seria

„Bazele matematice ale cercetării operaționale"

Au apărut :

– Maliță, M., Zidăroiu, C.
„Matematica organizării"

– Kaufmann, A., Desbazeille, G.
„Metoda drumului critic"

– Tomescu, I.
„Introducere în combinatorică"

– Teodorescu, N. ș.a.
„Metode ale cercetării operaționale în gestiunea întreprinderilor"

– Ghermeier, B. I.
„Bazele teoretice ale cercetării operaționale"

– Owen, G.
„Teoria jocurilor"

– Ackoff, R. L., Sasieni, M. W.
„Bazele cercetării operaționale"

– Lasdon, L. S.
„Teoria optimizării sistemelor mari"

– Zidăroiu, C.
„Programare dinamică discretă"

– Ermakov, S. M.
„Metoda Monte Carlo și probleme înrudite"

– Drăgan, I.
„Tehnici de bază în programarea liniară"

Lei 18

Editura tehnică